Optimization of Disaggregated Space Systems Using the Disaggregated Integral Systems Concept Optimization Technology Methodology

Katherine M. Wagner

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Aerospace Engineering

Jonathan T. Black, Chair
Cornel I. Sultan
Alan J. Michaels
Robert A. Canfield
Kevin K. Schroeder

June 2, 2020
Blacksburg, Virginia

Keywords: disaggregation, constellation design, DSS, optimization

Copyright 2020, Katherine M. Wagner

Optimization of Disaggregated Space Systems Using the Disaggregated Integral Systems Concept Optimization Technology Methodology

Katherine M. Wagner

(ABSTRACT)

This research describes the development and application of the Disaggregated Integral Systems Concept Optimization Technology (DISCO-Tech) methodology. DISCO-Tech is a modular space system design tool that focuses on the optimization of disaggregated and non-traditional space systems. It uses a variable-length genetic algorithm to simultaneously optimize orbital parameters, payload parameters, and payload distribution for space systems. The solutions produced by the genetic algorithm are evaluated using cost estimation, coverage analysis, and spacecraft sizing modules. A set of validation cases are presented. DISCO-Tech is then applied to three representative space mission design problems. The first problem is the design of a resilient rideshare-manifested fire detection system. This analysis uses a novel framework for evaluating constellation resilience to threats using mixed integer linear programming. A solution is identified where revisit times of under four hours are achievable for $10.5 million, one quarter of the cost of a system manifested using dedicated launches. The second problem applies the same resilience techniques to the design of an expanded GPS monitor station network. Nine additional monitor stations are identified that allow the network to continuously monitor the GPS satellites even when five of the monitor stations are inoperable. The third problem is the design of a formation of satellites for performing sea surface height detection using interferometric synthetic aperture radar techniques. A solution is chosen that meets the performance requirements of an upcoming monolithic system at 70% of the cost of the monolithic system.

Optimization of Disaggregated Space Systems Using the Disaggregated Integral Systems Concept Optimization Technology Methodology

Katherine M. Wagner

(GENERAL AUDIENCE ABSTRACT)

Civilians, businesses, and the government all rely on space-based resources for their daily operations. For example, the signal provided by GPS satellites is used by drivers, commercial pilots, soldiers, and more. Communications satellites provide phone and internet to users in remote areas. Weather satellites provide short-term forecasting and measure climate change. Because of the importance of these and other space systems, it is necessary that they are designed in an efficient, reliable, and cost-effective manner. The Disaggregated Integral Systems Concept Optimization Technology (DISCO-Tech) is introduced as a means of designing these space systems. DISCO-Tech optimizes various aspects of the space mission, including the number of satellites needed to complete the mission, the location of the satellites, and the sensors that each satellite needs to accomplish its mission. This dissertation describes how DISCO-Tech works, then applies DISCO-Tech to several example missions. The first mission uses satellites to monitor forest fires in California. In order to reduce the cost of this mission, the satellites share launch vehicles with satellites from other, unrelated missions. Next, DISCO-Tech is used to choose the placement of new ground stations for GPS satellites. Because GPS is an important asset, this study also assesses the performance of the network of ground stations when some of the stations are inoperable. Finally, DISCO-Tech is used to design a group of satellites that measure sea level, since sea level is important for climatology research. A design is presented for a group of satellites that perform these measurements at a lower cost than a planned mission that uses a single satellite.

Dedication

I would like to thank my husband Kurt first and foremost. His continual support and love throughout this daunting process has kept me sane, and I truly could not imagine getting through all of this without him. I would also like to thank my mother, Susan Mott, and my father, Bob Mott. My mother has been an inspiration to me and a source of love and encouragement for my entire life. My father taught me that I can do any task to which I set my mind, and can always make me smile whenever I am upset or stressed. I am so fortunate to have such wonderful parents. Finally, I want to thank the rest of my family. I would not be where I am today without all of them.

Acknowledgments

I would like to thank my advisor, Dr. Jonathan Black, for his guidance and support throughout my graduate research. He steered my research while giving me enough space to pursue the topics I found most interesting. I am also deeply appreciative toward Dr. Kevin Schroeder, who acted as a mentor and offered a significant amount of advice and support over the past couple of years. I am also grateful for the help of Dr. Cornel Sultan, Dr. Alan Michaels, and Dr. Bob Canfield. I appreciate their willingness to serve on my committee, their advice, and their feedback. I would like to thank Tom Krauss for sharing his radar expertise. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1651272. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Support for this work was also provided by the Virginia Tech Institute for Critical Technology and Applied Science, the Ted and Karyn Hume Center for National Security and Technology, and the Virginia Space Grant Consortium.

Contents

List of Figures

List of Tables

List of Abbreviations

Δv Change in velocity

ϵNSGA-II ϵ Nondominated Sorting Genetic Algorithm II

λ Wavelength

$\lfloor \cdot \rceil$ Rounding operator

$\mathbb{R}$ Set of real numbers

$\mathbb{Z}$ Set of integers

μ Standard gravitational parameter

ν True anomaly

Ω Longitude of the ascending node

Ω Right ascension of the ascending node

ω Argument of periapsis

τ Pulse width

E_b/N_0 Energy per bit to noise power spectral density ratio

J_2 Second degree zonal harmonic coefficient

r_E Planet radius

T_S Satellite nodal period

a	Semimajor axis
D	Scenario duration
E	Eccentric anomaly
e	Eccentricity
f	Frequency
h	Specific angular momentum
i	Inclination
k	Wavenumber, $2\pi/\lambda$
M	Mean anomaly
n	Generic integer
n	Mean motion
p	Semiparameter
R	Radius
r	Radius
u	Argument of latitude
AMR	Advanced microwave radiometer
AoP	Argument of periapsis
BER	Bit error rate
BW	Bandwidth

CERs Cost Estimating Relationships

CNES Centre national d'études spatiales

COTS Commercial off-the-shelf

DARPA Defense Advanced Research Projects Agency

DEM Digital Elevation Model

DISCO Disaggregated Integral Systems Concept Optimization

DISCO-Tech Disaggregated Integral Systems Concept Optimization Technology

DLR Deutsches Zentrum für Luft-und Raumfahrt (German Aerospace Center)

DOP Dilution of precision

DV Decision variable

EIRP Effective isotropic radiated power

ER Expansion rate

FOV Field of view

FSS Federated Satellite Systems

FY2020 Fiscal Year 2020

GA Genetic Algorithm

GNSS Global Navigation Satellite System

GPS Global Positioning System

GRIPS Genetic Resources for Innovation and Problem Solving

GSD Ground sample distance

GT-FAST Georgia Tech F6 Architecture Synthesis Tool

GVE Gauss's Variational Equations

HPBW Half-power beamwidth

InSAR Interferometric Synthetic Aperture Radar

KaRIn Ka-band radar interferometer

LAN Longitude of the ascending node

LCR Length change rate

LEO Low Earth orbit

LP Linear program

LVLH Local vertical, local horizontal

MCDM Multi-criteria decision making

MEOE Modified equinoctial orbital elements

MILP Mixed integer linear programming

MOMINLP Multi-objective mixed integer nonlinear program

MTTA Mean time to access

MV Mixed variable

NASA National Aeronautics and Space Administration

NICM NASA Instrument Cost Model

NSGA-II Nondominated Sorting Genetic Algorithm II

ODE Ordinary differential equation

PNT Positioning, Navigation, and Timing

POD Precision orbit determination

PRF Pulse repetition frequency

RAAN Right ascension of the ascending node

SAR Synthetic Aperture Radar

SNR Signal-to-noise ratio

SODA Systems Operational Dependency Analysis

SSCM Small Satellite Cost Model

SSH Sea surface height

STK Systems Tool Kit

SWH Significant wave height

SWOT Surface Water and Ocean Topography

TAG Time average gap

TanDEM-X TerraSAR-X add-on for digital elevation measurement

TAT-C Tradespace Analysis Tool for Constellations

TDX TerraSAR-X add-on for digital elevation measurement

TLE Two-line element set

TRL Technology readiness level

TSX TerraSAR-X

TT&C Telemetry, tracking, and command

TWTA Travelling wave tube amplifier

UHF Ultra High Frequency

USCM Unmanned Space Vehicle Cost Model

VASSAR Value Assessment for System Architectures using Rules

VLEO Very low Earth orbit

Chapter 1

Introduction

The overarching objective of this research is to develop a methodology for disaggregation optimization that is broadly applicable and addresses gaps in capability for constellation and spacecraft design. This broadly applicable methodology is described in detail and applied to a variety of space mission design problems.

This chapter introduces the motivation and goals for this research effort. Chapter 2 surveys existing methodologies for space mission and constellation design with a focus on methodologies for optimizing disaggregated space systems. Gaps in capability for disaggregated space systems are identified. Chapter 3 describes the development and capabilities of the Disaggregated Integral Systems Concept Optimization Technology (DISCO-Tech), a constellation design tool that emphasizes disaggregation, nontraditional constellations, and resilience. A variable-length genetic algorithm capable of simultaneously optimizing payload distribution, payload parameters, and orbital parameters is presented. The sizing methodology used by DISCO-Tech is described. Common DISCO-Tech modules such as coverage, cost, and reachability are discussed. Chapter 4 explores some test cases to validate the DISCO-Tech methodology. The first case assesses the ability of DISCO-Tech to develop a reasonable solution for a Mars-orbiting communications, navigation, and data relay constellation. DISCO-Tech identifies a homogeneous solution of 45 satellites with a total cost of $2.9 billion. The second case compares the performance of a variable-length genetic algorithm and a flagged genetic algorithm for spacing orbital planes to provide coverage over a target

region. The variable-length genetic algorithm outperforms the flagged genetic algorithm and produces a nondominated set of solutions that is similar to the set produced using Walker constellations. The final validation case uses DISCO-Tech to perform a payload assignment problem for a rules-based fractionation problem. The results are compared to the results produced by an exhaustive search of all combinations of payloads. The genetic algorithm finds the true set of nondominated results in 3% of the number of function evaluations used by the exhaustive search. Chapter 5 uses DISCO-Tech to design resilient systems for two scenarios. The first scenario uses rideshare opportunities to manifest a constellation for fire detection. A rideshare solution providing discontinuous coverage is identified that provides similar capability to a two-plane, four-satellite Walker constellation at about 25% of the cost. The second scenario optimizes the placement of new GPS monitor stations, considering the performance of the monitor station network when some stations are inoperable. An additional nine station locations are identified that allow the GPS monitor station network to provide continuous coverage even when five monitor stations are inoperable. Chapter 6 uses DISCO-Tech to design a group of satellites flying in formation that use synthetic aperture radar interferometry to measure sea surface height. Trade-offs between different performance objectives, such as height error and cost, are assessed. A solution is identified that uses two heterogeneous formations of ten satellites each. The solution offers a 30% cost savings compared to a monolithic solution. Chapter 7 summarizes the contributions, results, and future work for the research.

The DISCO-Tech logo is shown in Fig. 1.1.

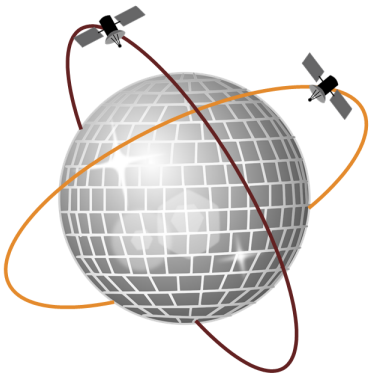


Figure 1.1: DISCO-Tech logo.

1.1 Motivation

Constellation design is a key aspect of any space mission. Historically, the needs of the payload have dictated the orbit selection process; the payload was designed first, then the orbit or orbits were chosen based on limitations imposed by the payload on altitude, lighting conditions, inclination, and launch vehicle selection. The orbits generally fell into one of several categories: sun-synchronous orbits, Molniya orbits, geosynchronous orbits, and Walker constellations.

Although this process has advantages in simplicity, heritage, and optimality for specific types of missions, it sacrifices generality. It also sacrifices the ability to simultaneously optimize the payload parameters and the orbital parameters. Furthermore, the methods described above are appropriate mainly for constellations in which all the satellites in the constellation are identical, termed *homogeneous constellations*. *Heterogeneous constellations*, in which the satellites comprising the constellation have varying configurations and payloads, lead to additional design challenges. For these reasons, constellation design in recent years has turned to metaheuristic optimization techniques such as genetic algorithms. Modern tools

and methodologies have been developed that each address certain aspects of the constellation design problem but lack the broad applicability needed to assess the design of heterogeneous systems, including disaggregated space systems.

Several trends in space missions have highlighted the potential of and need for disaggregated space systems. The first trend of importance is the increase in low-cost, low-mass satellites. Over the past couple of decades, there has been an increase in small satellite, microsatellite, and nanosatellite missions. According to a study published by Bryce Space and Technology, the number of nanosatellites (1–10 kg), picosatellites (0.1–1 kg), and femtosatellites (0.01–0.1 kg) has increased tenfold between 2012 and 2017. In 2017, there were 300 satellites launched with masses less than 10 kg. There are a few contributing factors that have led to this boom in small space missions. The first trend is the gradual reduction in electronics size that occurs in most areas of electronics. The second contributing factor is the creation of the CubeSat standard [72]. The third, related trend is the increasing availability of commercial off-the-shelf (COTS) satellite components, particularly for CubeSats. The final factor is the existence of rideshare programs, in which small satellites are launched either as a secondary payload with a larger, more expensive primary payload or with a large number of other low-mass satellites. Disaggregation can take advantage of the developments in low-mass satellites by using these satellites to replace more massive monolithic satellites.

The second trend that has increased interest in disaggregation is the rising mass of space systems. Mass is a chief indicator of complexity, and complex systems are more likely to experience cost overrun, schedule overrun, and program cancellation [14]. A disaggregated system has the potential to replace a massive and complex single satellite with multiple less massive, less complex vehicles. This strategy can not only decrease cost but can also decrease programmatic risk and ensure that some capability is maintained even if an architecture is only partially deployed [152].

The final trend that has highlighted a need for resilience is increasing global reliance on space systems. As space-based assets become increasingly vital to military and civilian operations, the need to protect those capabilities from disruption in service increases. Disruption in service can come either from intentional action or from system failure. Early space systems were not designed to be resilient to intentional disruptions due to the political climate of the Cold War. It was assumed during that time that actions taken against space-based assets would be seen as acts of war and were therefore unlikely to be attempted due to "mutually assured destruction." The current geopolitical climate has changed sufficiently such that the inherent safety of space systems is no longer guaranteed [30]. Additionally, the increasing commoditization of space has the potential to move some critical assets outside of governmental control and into the commercial sector. These assets may lack the inherent protection of government backing while remaining critical to military and civil operations. In addition to intentional actions, spacecraft are subject to failure from a variety of other sources, including launch failures, premature component failures, and collisions. These vulnerabilities highlight the fact that a single spacecraft is ill-equipped to respond to an anomaly, putting the entire mission at risk. For this reason, Air Force Space Command has identified disaggregation as a potential strategy for increasing the resilience of space systems [30].

The Air Force defines disaggregation as "the dispersion of space-based missions, functions or sensors across multiple systems spanning one or more orbital plane, platform, host or domain," [30]. Disaggregation can be broken down into five subcategories:

1. *Multi-orbit disaggregation:* Multi-orbit disaggregation is the process of using satellites spread across multiple orbital planes to complete a mission. This method of disaggregation is commonly employed not only for its resiliency benefits but also its coverage benefits. Multiple planes at the same altitude and inclination can be used to provide continuous global coverage or to reduce revisit times. Orbits at different altitudes

can be used to provide a mix of good coverage through higher altitudes and increased sensor performance at low altitudes.

2. *Multi-domain disaggregation:* Multi-domain disaggregation involves combining space-based capabilities with sensors in other domains, such as drones, high-altitude balloons, ground vehicles, ground sensors, ships, and submarines. Non-space assets are generally easier to deploy and can provide higher resolution measurements due to close proximity than can satellites. However, satellites offer greater coverage than these alternate solutions. By combining capabilities across domains, it is possible to leverage the unique advantages of each domain.

3. *Functional disaggregation:* Functional disaggregation is the process of dividing a payload that could have been hosted on a single, monolithic satellite among multiple smaller satellites. For example, a mission requiring a radar, a mid-wave IR sensor, and an optical camera could potentially use three satellites (one per sensor) instead of one satellite. This process reduces the complexity of each individual satellite compared to the monolith. It also increases the resiliency of the mission, as the loss of a single spacecraft does not necessarily ruin the entire mission. Redundant sensors could also be included on different spacecraft to further increase resilience at the expense of increased complexity and cost.

4. *Fractionation:* Fractionation is a more extensive version of functional disaggregation. Whereas functional disaggregation involves distributing sensors across satellites, fractionation involves dividing the mission among a number of modules. The modules interact wirelessly with one another. While a sensor is one type of module, modules can also include various spacecraft subsystems. Unlike a constellation formed through functional disaggregation, some of the individual components of a fractionated constellation are generally unable to function as a satellite in the absence of other modules.

Communications capabilities are the primary candidate for fractionation. In the fractionated constellation, only some of the modules may be capable of communicating with the ground. Other modules receive commands from and pass data through the communicating modules. Other capabilities that have been identified as candidates for fractionation are on-board computation and data processing, power generation (with power passed to other modules via lasers), and attitude and orbit determination [124].

5. *Hosted payloads:* Hosted payload disaggregation is a unique method of disaggregation that fields the sensors needed for a mission by adding those sensors onto other spacecraft. In this method, the host spacecraft has its own mission, often unrelated to the hosted payload mission. The hosted payload shares resources with the primary payload, such as power, data downlinking, and computation resources. Hosted payloads reduce the costs required to deploy a sensor. However, the hosted payloads are constrained in size, power, and orbit by the availability of hosting satellites.

Disaggregated constellations can therefore take many forms. It is necessary to be able to compare various disaggregated constellations and to benchmark those constellations against monolithic satellites capable of performing the same missions. The current state of the art lacks the ability to dynamically optimize constellations of various disaggregation levels. This capability gap is the motivation for the creation of DISCO-Tech.

1.1.1 Research Questions

This project seeks to answer several questions with regards to disaggregated system modelling and applicability.

1. How can multiple levels of disaggregation be easily represented and optimized?

Hypothesis: A variable-length genetic algorithm can be used to optimize disaggregated space systems.

2. How can resilience be assessed in a tractable and comprehensive manner?

   Hypothesis: Traditional revisit metrics and linear programming can be used to fully explore the threat space and determine disadvantaged system performance.

3. What are the trade-offs between cost, performance, resilience, and complexity for disaggregated and non-traditional systems?

   Hypothesis: Disaggregated systems can trade vehicle complexity (heavy, complex single satellite) for constellation complexity (many satellites but low individual complexity). Disaggregation will increase resilience compared to a monolithic system. Ad hoc constellations will offer decreased cost and increased resilience for small satellite missions up to some limit on performance. For coverage-based analyses, increasing disaggregation will improve performance. Cost can be traded for either performance or resilience.

1.1.2 Research Objectives

In order to develop a broadly applicable methodology for designing disaggregated space systems and to answer the research questions defined in the previous section, several subobjectives must be accomplished, each one addressing a separate aspect of the research. The first subobjective is to explore the viability of using a variable-length genetic algorithm to concisely define and optimize a disaggregation-eligible constellation design problem such that it is not necessary to define specific levels of disaggregation a priori and to compare the variable-length genetic algorithm to alternatives like a flagged genetic algorithm. This subobjective addresses research question #1 and is discussed in section 3.2. The second subobjective is to develop a mixed integer linear programming formulation for identifying

the worst-case loss of satellites for a broad range of constellation configurations, particularly asymmetric constellations with many assets. This subobjective addresses research question #2 and is addressed in subsubsections 5.2.2 and 5.3.2 and in Appendix B. The third subobjective is to leverage existing cost and mass models in addition to aggregated program data to provide mass and lifecycle cost estimates based on design parameters for disaggregated systems. This subobjective is used in answering research question #3 and in forming the overall DISCO-Tech framework. It is discussed in sections 3.4 and 3.6. The fourth subobjective is to develop custom, application-specific objective and constraint functions and integrate them into the DISCO-Tech framework. This subobjective is needed to address research question #3. Custom objectives and constraints are used in the scenario in Chapter 6. The final subobjective is to be able to represent the reconfiguration of the constellation, either to move satellites from their launch configuration to a final configuration or to change configuration based on knowledge of changing mission needs over time. This subobjective is necessary for addressing research question #3 for reconfigurable constellations, such as the rideshare constellation in section 5.2.

When combined with an object-oriented code base that easily allows analysts to define new disaggregation optimization problems, these subobjectives combine to form a tool that provides a unique and powerful capability. The broad applicability and ease of use of DISCO-Tech make it an enabling technology in a world in which resilience considerations are increasingly important.

1.1.3 Mission selection

Three principal applications are chosen for exploration as part of this research project. The first is a rideshare-manifested imaging constellation for the purpose of fire detection. The

second applies the DISCO-Tech methodology to the problem of determining ground station placement for an expansion of the Global Positioning System (GPS) ground station network. The third is a disaggregated constellation performing sea surface height measurements using interferometric synthetic aperture radar (InSAR). Additionally, three secondary applications are explored for validation purposes: The problem of designing a communications and positioning, navigation, and timing (PNT) constellation around Mars for early human habitation; the optimization of a multi-plane satellite constellation for continuous communications coverage; and the optimization of a fractionated space system for a set of payloads.

Rideshare-manifested imaging constellation

Developments in small satellite technology have lowered the barrier to entry to space, both financially and technologically. As a result, space missions are now a feasible solution for entities with limited funding, such as universities and startups. Additionally, spacecraft made from commercial off-the-shelf parts can be constructed on shorter timelines than was previously possible, increasing responsiveness and rapid deployment of space systems. This increased responsiveness can be leveraged to launch systems in response to disasters, such as hurricanes or fires. To explore the performance achievable with low-cost satellite constellations, the problem of detecting nascent fires in the State of California is examined.

Estimates place the total damage caused by California wildfires in 2018 at $400 billion, with $1 billion spent by the California fire department [117]. Fire detection is typically accomplished using manned fire towers, but visible detection of small fires is difficult during the day. There are space-based assets whose imagery could be used to detect fires, but such systems do not have sufficient revisit rates to detect all early fires. This research examines the ability of rideshare-manifested nanosatellites, constructed from COTS components, to

image, detect, and report wildfire development in California. These results are compared to Walker constellations comprised of the same satellites. Hosted payloads are not considered in this scenario because the spacecraft will be required to slew to cover the area of interest. It is assumed that a primary payload would not permit such motion.

This scenario addresses research question # 2 through the examination of degradation for an asymmetric constellation. It also addresses research question #3 by comparing reconfigurable rideshare-based constellations to Walker constellations.

GPS ground station expansion

GPS is a key asset for civilian, commercial, and economic applications. To maintain accurate knowledge of their positions, the GPS satellites rely on monitor stations located around the globe. The stations continuously collect data transmitted by the GPS satellites. That data is combined with atomic clock measurements generated on-site, then passed to the GPS Master Control Station at the U.S. Air Force's 2nd Space Operations Squadron (2SOPS). These measurements provide the information needed to estimate the ephemerides of the GPS satellites and to correct clock errors on board the satellites. The current configuration of ground stations is such that each GPS satellite is always within view of at least three ground stations. Further expansion of the GPS ground station network would increase the accuracy of the measurements. More importantly, it would provide redundancy in the event that one or more stations suffered a temporary loss of service. This scenario examines the optimal placement of additional ground stations. It also considers the worst-case performance when multiple ground stations are lost. This scenario addresses research question #2 by assessing ground station resilience to threats.

InSAR sea surface height monitoring constellation

Cost-benefit studies on disaggregation have thus far been focused mainly in a few fields: weather [70, 95, 97, 151], fire detection [50, 150] and PNT [63, 69, 82, 95]. Additional research efforts are needed to study the applicability of disaggregation to other mission types. One field of study that can potentially benefit from disaggregation is climate monitoring. The 2007 National Research Council Decadal Survey [32] identified sea surface height (SSH) and surface water levels as two important metrics in global climatology. SSH, which is typically measured using nadir-pointing radar altimeters, provides information about ocean currents and weather patterns. A desire for finer temporal and spatial resolution for SSH measurements resulted in the planned National Aeronautics and Space Administration (NASA)/Centre national d'études spatiales (CNES) Surface Water and Ocean Topography (SWOT) mission, which will measure both SSH and surface water height using interferometric synthetic aperture radar (InSAR). However, SWOT's novel instrumentation is causing schedule delays and cost overruns. The estimated life cycle cost for the program is $1.14 billion, but technical issues in the primary payload and cost modelling deficiencies led an audit to suggest that the allocated funds may be insufficient [5]. This analysis examines whether a disaggregated constellation could be used to perform wide-swath InSAR and accomplish the goals of the SWOT mission. The project focuses on SSH monitoring, but the methodology is designed for flexibility and can be extended to surface water monitoring.

This scenario addresses research question #3 by examining different levels of disaggregation and different disaggregation types and comparing the disaggregated performance to that of the baseline SWOT program #3.

Validation cases

Three problems are presented as validation cases for the DISCO-Tech methodology. The first case is the design of a Mars-orbiting communications, PNT, and data relay constellation. Interest in sending manned missions to Mars has increased in the last decade. Indeed, various organizations have declared their intent to perform a manned Mars mission within the next two decades [20, 62]. For manned missions to Mars to become a reality, it will be necessary to have communications infrastructure available to support both planetary and interplanetary communications. The goal of this application is to produce an initial design of a Mars-orbiting communications constellation that can be reevaluated as mission requirements and available technologies change. This analysis assumes a Walker constellation and optimizes the transmitter properties and payload distribution. This scenario addresses research question #1 by confirming the ability of a genetic algorithm to simultaneously optimize orbital and payload parameters. The second validation case optimizes the distribution of orbital planes for the case of providing coverage between 60°S and 60°N. A genetic algorithm is used to vary the number of orbital planes, the number of satellites per plane, and the relative position of the planes. The results are compared to the performance achievable with a Walker constellation. This scenario addresses research question #1 by evaluating the ability of a genetic algorithm to dynamically vary the number of satellites and planes in a constellation. The final scenario optimizes the distribution of a fixed set of payloads among modules in a fractionated system. A rules-based approach is used to determine resilience and cost in this scenario. The results are compared to the solutions found when performing a full factorial search of the design space. This scenario addresses research question #1 by assessing the ability of a genetic algorithm to dynamically vary the number of different types of systems and the payload distribution in a heterogeneous cluster of modules.

Chapter 2

Review of Literature

This chapter reviews the existing body of work in the field of disaggregated system design and discusses the areas in which DISCO-Tech will expand the current body of research. First, methodologies for the design of homogeneous constellations are reviewed. Analytically derived constellation geometries, such as Walker constellations, are discussed in the first section. DISCO-Tech leverages these constellation geometries to characterize the orbits for various missions. This section also discusses the use of genetic algorithms and other metaheuristic techniques in constellation design, which are used as the basis for the custom genetic algorithm developed for DISCO-Tech. The next section discusses existing tools for the optimization and evaluation of disaggregated space systems. The capabilities and limitations of existing methodologies are discussed. Gaps in the capabilities are identified that DISCO-Tech will address. The third section discusses missions to which disaggregation has been applied, either in theory or in practice. DISCO-Tech provides a new formulation for fire detection, a scenario that has been assessed using disaggregation techniques in previous studies. DISCO-Tech also examines using a disaggregated architecture for a sea surface height monitoring constellations, an area in which no cost-benefit analysis has been performed for disaggregated systems. Finally, disaggregated systems in other domains and the possibility of multi-domain architectures are discussed. DISCO-Tech is applied to the design of a network of ground stations and can easily be adapted to perform multi-domain analysis.

2.1 Constellation design methodologies

2.1.1 Walker delta constellations

A commonly used constellation is the Walker Delta constellation. A Walker Delta constellation consists of multiple circular orbits evenly distributed in right ascension of the ascending node (RAAN) about Earth's axis. Satellites within an orbital plane are evenly spaced in true anomaly. Satellites are positioned relative to adjacent planes according to a phasing parameter. All orbits have the same semimajor axis, inclination, and zero eccentricity. This symmetric constellation is often used for scenarios in which the region of interest is the entire planet or some latitude band centered on the equator. Walker showed that this configuration can provide global coverage with as few as five satellites [159]. Walker constellations are used in the DISCO-Tech methodology as the default configuration for a constellation providing global coverage. DISCO-Tech can optimize the parameters used by the Walker constellation such as the reference orbit, the number of planes, the satellites per plane, and the phasing parameter. The ability of DISCO-Tech to generate solutions that perform similarly to Walker constellations is assessed in Section 4.3. Walker constellations also form the basis of the orbit design in Section 4.2.

2.1.2 Flower constellations

Flower constellations are a relatively recent extension of Walker delta constellations and were introduced by Mortari, Wilkins, and Bruccoleri in 2004. They extend the Walker formulation to elliptical orbits. Flower constellations use repeating ground tracks to force multiple satellites across multiple planes to follow the same ground track. All orbits in

the flower constellation have the same inclination, eccentricity, argument of periapsis, and semimajor axis [113]. Lattice flower constellations are the subset of flower constellations for which a maximal number of satellites are used and for which the minimal description of such a formulation is employed [39, 40]. DISCO-Tech can be used to optimize the parameters for a flower constellation.

2.1.3 Metaheuristic techniques

Metaheuristic techniques are frequently used to explore the constellation design space. These techniques are popular due to their generality, their ability to explore non-convex spaces, and their ability to solve multi-objective optimization problems. Genetic algorithms (GA) in particular have been applied in many cases to the problem of constellation design [6, 23, 55, 56]. They have been shown to be superior or comparable to Walker constellation design techniques [126]. One specific tool for constellation optimization is the Genetic Resources for Innovation and Problem Solving (GRIPS) tool from the Aerospace Corporation, which has been applied to a range of constellation design problems [54, 61, 166]. Hitomi used a variable-length genetic algorithm to vary the number of satellites in a constellation design problem [73]. GAs have also been used for hardware selection for spacecraft [105]. Other metaheuristic techniques such as simulated annealing have successfully been applied to constellation design [34]. A description of the custom genetic algorithm used by DISCO-Tech is given in Section 3.2.

2.2 Disaggregated constellation design tools

There are a handful of constellation design tools focused on various types of disaggregated space systems. The tools described in this subsection represent the state of the art for disaggregated constellation design. Since this type of system is DISCO-Tech's focus, the gaps identified here are used to develop research questions for this project.

The first tool of interest is the Tradespace Analysis Tool for Constellations (TAT-C). TAT-C is a tool for the evaluation of disaggregated architectures being developed at NASA Goddard Space Flight Center. TAT-C's search mechanism is a full factorial exploration of a discretized tradespace. The user defines ranges for orbital parameters, frequencies, and fields of view. TAT-C calculates coverage metrics, signal-to-noise ratios (SNR), and resolution. The resulting solutions are sorted into nondominated sets and reported. Cost Estimating Relationships (CERs) from popular models like the Small Satellite Cost Model (SSCM), Unmanned Space Vehicle Cost Model (USCM), and NASA Instrument Cost Model (NICM) are combined to estimate system cost. It also calculates probabilistic cost distributions based on expected failure rates. It performs launch manifesting by assuming that all satellites are launched simultaneously on the same type of launch vehicle. Although TAT-C documentation suggests that it is capable of evaluating heterogeneous constellations, no studies have been published that use this feature. The heterogeneous constellations would be generated through a full enumeration of all possible combinations of payloads and would not consider redundancy. In the published studies, TAT-C optimizes for orbital parameters and payload parameters only [59, 110, 118, 119].

The Value Assessment for System Architectures using Rules (VASSAR) methodology also assesses performance of disaggregated architectures. Solutions are constructed using a predefined set of orbits and payloads. Each solution describes a mapping of payloads to or-

bits. VASSAR then uses a rules-based system to estimate performance. Predefined relationships such as the effect of orbit choice on revisit time, resolution, and access; effect of payload assignments on required downlink rate; and synergies between payloads due to cross-registration with co-located sensors are used to assess performance. Resilience is assessed for single satellite loss. VASSAR uses a domain-dependent metaheuristic algorithm to explore the design space. This algorithm includes mutation and crossover operations used by genetic algorithms but uses additional heuristics that drive solutions toward known preferred configurations (e.g. synergistic payload combinations and orbits that are preferred by specific payloads). VASSAR does not optimize for payload parameters [23, 74, 133, 134, 157].

The Disaggregated Integral Systems Concept Optimization (DISCO) tool is the predecessor to DISCO-Tech. DISCO was developed at the Air Force Institute of Technology to enable comparison between disaggregated systems and monolithic systems. It uses a single objective genetic algorithm to explore the tradespace and simultaneously optimizes payload parameters, such as aperture size for optical systems, and orbital parameters. It uses common mass and cost models like SSCM, USCM, and NICM to estimate the lifecycle cost for a given level of disaggregation. It integrates with Systems Tool Kit (STK) to perform propagation and coverage analyses or uses analytical equations to approximate coverage [152]. DISCO does not dynamically optimize the payload distribution but compares a predefined disaggregation scheme to a monolithic system.

The Georgia Tech F6 Architecture Synthesis Tool (GT-FAST) is designed to evaluate the cost and performance of fractionated systems. GT-FAST performs a combinatorial analysis of fractionated architectures. Its inputs are fractionated configuration, launch vehicle manifest, orbit, design life, payload mass, payload power, payload performance requirement, and specific impulse. The tool outputs estimated system mass, power, and cost budgets both for the overall cluster and for each module. Additional output metrics include the number of

launches necessary to manifest the constellation, the average cost of replacing a module, and resilience as measured by the average percentage of capability lost with a random module failure. It does not explore the orbital design space or payload parameters [96].

Work by DeLaurentis also focuses on fractionated architectures. One paper compares the complexity of modular, fractionated, and monolithic systems by examining the number of nodes and links in the system using a remote sensing mission as an example [146]. Another examines the trade-offs between mass, cost, and other parameters for fractionated systems [145]. The Systems Operational Dependency Analysis (SODA) methodology analyzes the impact of dependencies between systems on performance by characterizing the strength, criticality, and interactions of the systems and quantifying operability after subsystem loss [66]. A recent paper uses machine learning to synthesize system-of-systems architectures and identify features common to architectures that exhibit strong performance. New architectures are then generated based on those features. Components are selected from a predefined list of options, and orbits are selected from one of several orbital regimes. This methodology is applied to the case of designing a set of fractionated architectures. The optimization objectives are mass, robustness, resilience, cost of servicing, serviceability, and time without servicing [67]. This research does not optimize payload parameters.

In addition to tools described above that attempt a tradespace exploration for various disaggregation schemes, there are tools that assess the cost and mass of predefined architectures. O'Neill and Weigel produced a cost/mass modelling methodology for fractionated systems that considers three levels of fractionation: monolithic, shared attitude/position determination and communications, and shared attitude/position determination, communications, and power. The analysis includes various subfunctions for estimating optical instrument diameter and data rate based on resolution; estimating computer/command system size; antenna sizing based on data rates and time that ground stations are in view; IMU/star tracker

selection based on pointing tolerances; solar array/battery sizing based on power needs and amount of light; reaction wheels and propellant sizing based on disturbance torque, pointing tolerances, and slewing needs; and thermal subsystem sizing based on heat dissipation and acceptable temperature ranges. The output of the tool is an estimation of the mass, power, and size of the modules [124]. Another tool estimates the expected cost of fractionated systems using a Weibull distribution to represent the expected lifetime of the modules and a Monte Carlo analysis to explore and roll up various possible instances [122].

La Tour's research involved applying system dynamics modelling techniques to the problem of constellation design. The system dynamics methodology involves tuning a model based on known results such that the tuned model will be valid over a range of inputs. This technique provides a multi-generational assessment of costs for a predefined disaggregated architecture. There is no focus on orbit optimization, and systems are instead assumed to be flown in clusters around the positions of spacecraft in the benchmark programs [69, 70, 95].

Table 2.1 summarizes the capabilities of existing disaggregation methodologies and of DISCO-Tech. Based on Table 2.1, it is apparent that there are no existing tools for simultaneously optimizing payload parameters, payload distribution, and orbital parameters. Simultaneous optimization across all areas will lead to designs that more completely integrate satellite design with orbit design, as both are dependent upon the disaggregation scheme. DISCO-Tech will fill this gap in capability using the techniques discussed in Chapter 3. Additionally, existing methodologies do not focus on the performance of a system in the event of one or more spacecraft failures, a property called resilience. Although GT-FAST examines resilience for fractionated systems, the same analysis cannot be easily extended to multi-orbit systems that must provide coverage to ground points. VASSAR's methodology for resilience after a single spacecraft loss is also not easily extensible due to the growth in problem size as the number of spacecraft and the number of spacecraft losses increase. A method for assessing

resilience in coverage-based analyses is discussed in Chapter 5.

Table 2.1: Lit review table.

	GT-FAST	TAT-C	VASSAR	DISCO	DISCO-Tech
Author	Lafleur	Le Moigne et al.	Selva et al.	Thompson	Wagner
References	[96]	[59, 110, 118, 119]	[74, 133, 134, 157]	[28, 150, 151, 152]	[114, 115, 116, 158]
Disaggregation types	FR	MO/FD	MO/FD/HP	MO/FD	FR/MO/FD/ HP/MD
Orbit optimization		X		X	X
Payload param. optim.		X		X	X
Disaggregation optim.	X		X		X
Resilience	X		X		X
Redundancy			X		X
Maneuvering		X			X
Launch manifest				X	X

FR = Fractionation, MO = Multi-orbit disaggregation, HP = Hosted payload, FD = Functional disaggregation, MD = Multi-domain disaggregation

2.3 Disaggregated constellation applications

TAT-C has been applied to the problem of designing a multi-satellite LandSat constellation for imaging and a wide-angle radiometer mission using either a Walker constellation or rideshare opportunities provided by the Planet Labs constellation, though the latter case used randomly selected slots rather than optimized selections [118]. DISCO has been applied to a variety of problems including space weather disaggregation in response to the cancellation of

the NPOESS program [151], fire detection [150], and launch manifestation of heterogeneous space systems [28]. La Tour's system dynamics technique was used to compare costs for the cancelled NPOESS program to a disaggregated configuration in which weather sensors were placed on one satellite and climate sensors were placed on another. Various levels of disaggregation were also compared to JPSS [70]. The system dynamics technique was also applied to the problem of GPS disaggregation [69].

Fractionated architectures have been assessed for imaging [124] and generic Earth observation missions [25, 26]. Ref. 45 provides an analysis of the cost risk associated with fractionated architectures versus monoliths using a Markov chain analysis. In addition to the applications by specific methodologies outlined above, disaggregated architectures have been explored independently for CubeSat monitoring of water resources [31], fire detection [50], ice monitoring [156], and comet exploration [24]. A team at Stanford has published multiple papers on a functionally disaggregated occulter/telescope system. The proposed system is comprised of a microsatellite with an occulter and a 6U CubeSat with a telescope. The CubeSat is constructed mainly from COTS parts to reduce cost [88, 89, 92]. The same group has proposed small-scale disaggregated missions for gravitational measurements and special relativity applicability studies [38]. These missions are just a sample of studied disaggregated missions and not an exhaustive list.

An architecture called the Federated Satellite Systems (FSS) paradigm has been proposed that is a pseudo-hybrid of fractionated and hosted payload architectures. The concept behind FSS is that many unrelated systems will collaborate by sharing downlink bandwidth, storage, processing power, and instrument time. The concept is referred to as "cloud computing in space." Analyses have been published for the concept of a Global Navigation Satellite System (GNSS) using the Iridium satellites as resources in a federated system [63].

DISCO-Tech will focus on two space system applications. The first application is a fire

detection system. This application was chosen because the original DISCO research identified that fire detection could benefit from disaggregation. The analysis presented here differs from the original DISCO scenario as it uses rideshare opportunities and COTS parts to reduce cost. It also considers resilience as a key component of the optimization. The second application is a sea surface height monitoring system. This application was chosen because a planned monolithic system for measuring sea surface height is experiencing cost and schedule overruns. Disaggregation may provide a means of reducing programmatic risk and vehicle complexity for future sea surface height monitoring missions.

2.3.1 Funded disaggregated missions

The previous section highlights the rich body of research involving disaggregated systems. Disaggregated missions are not limited to conceptual designs and studies; many important space systems rely on various levels of disaggregation to complete their missions. Some missions, like GNSS, rely on distributed space systems due to the technical requirements for the mission and could not be accomplished with single satellite systems. Other systems use disaggregation to reduce risk, lessen complexity, or test future payloads. A handful of these missions are discussed herein.

Ref. 13 surveys existing distributed space systems (DSS), where a DSS is defined as "a system of multiple satellites designed to work in a coordinated fashion to perform a mission." The DSS concept could be considered a subset of disaggregation. Disaggregated systems employing multi-orbit disaggregation, functional disaggregation, or fractionation are analogous to distributed space systems. A hosted payload mission using one or more vehicles is also a DSS. GPS and GLONASS are referenced as GNSS DSSs. Remote sensing DSSs are also highlighted, including EOS (imaging, crop monitoring), COSMIC (climatology), and

THEMIS (magnetosphere measurements) [13].

An early proposed disaggregated mission was the TechSat-21 program, a concept in the late nineties and early 2000s. The TechSat-21 architecture was comprised of 35 small satellites, each with a radar and phased array antenna. The goal of the mission was to use the satellites in concert to produce synthetic aperture radar (SAR) imagery and to track ground vehicles. The satellites were to act as a sparse aperture, with each satellite illuminating the same area on the ground [106]. The program was ultimately cancelled due to unforeseen technical difficulties, cost overruns, and schedule overruns.

The TerraSAR-X/TanDEM-X system is an example of a distributed space system in which two satellites work together to perform SAR interferometry. The two satellites together form both bistatic and monostatic radars. Prior to TerraSAR-X, there were several demonstrations of multi-domain bistatic radar, generally with a space-based transmitter and an airborne or ground-based receiver [42].

The Sentinel satellites, which were launched as part of the ESA Copernicus program, are a disaggregated system of heterogeneous satellites. Sentinel-1A and Sentinel-1B were launched in 2014 and 2016 and perform radar imaging. Sentinel-2A and Sentinel-2B were launched in 2015 and 2017 and perform optical imaging. Sentinel-3A and Sentinel-3B were launched in 2016 and 2018 and use radar altimetry and radiometry to monitor the ocean. Sentinel-5P was launched in 2017 and performs atmospheric and air quality measurements. At least three more iterations of Sentinel satellites are planned. The data from the Sentinel satellites are combined to provide information about climate change, land monitoring, ocean monitoring, atmospheric monitoring, and more [135].

There have been a handful of government-sponsored hosted payloads flown thus far. One such example is the REACH program, a hosted payload dosimeter. The 1 kg payload was

placed on Iridium NEXT satellites for a total cost of $16M [103]. Others include GeoTRACE, which provided global observations of the Earth's limb and disk; Wide Area Augmentation System, which used an L-band payload to improve GPS performance for aircraft; Automatic Information System, a payload that forwarded ship transmissions to extend the range at which ships can be tracked; Internet Router In Space, a payload for demonstrating routing in space; Commercially Hosted Infrared Payload, an IR sensor hosted on a geostationary satellite for missile detection; and the Australian Defense Force Ultra High Frequency (UHF) payload, which provides UHF-band communications capabilities. An analysis of these programs estimated hosted payload costs at $6.6M + $0.175M/kg [9].

2.4 Multi-domain design

Similar efforts in industry and in other domains highlight interest in the concept of heterogeneous swarms of vehicles. In particular, several Defense Advanced Research Projects Agency (DARPA) projects are of particular interest. The DARPA Gremlins program, in development by Dynetics, aims to use lightweight, unmanned air systems as a basis for a variety of payloads [47]. The DARPA Hydra program seeks to extend U.S naval capabilities by augmenting manned Navy vessels with modular, unmanned, undersea payloads, describing a system that is not only heterogeneous but also multi-domain [123]. The DARPA Complex Adaptive System Composition And Design Environment (CASCADE) program calls for the development of a "unified view of system behavior" for complex, system-of-systems architectures, highlighting the need for tools that effectively manage such complex systems. The proposed research is capable of addressing these needs and providing a platform with broad applicability to multi-domain mission design problems. Additionally, space-based sensors can be combined with ground sensors to provide space situational awareness in multiple

domains. Simultaneous optimization of such a network is demonstrated in Ref. 29.

DISCO-Tech is applied to the optimization of GPS monitor station placement in Chapter 5. This application highlights the ability of DISCO-Tech to optimize missions that are not satellite missions. The DISCO-Tech architecture permits simultaneous optimization of vehicles and sensors across multiple domains and could be used to design such a mission in the future.

2.5 Summary

This section outlined existing constellation design strategies for both traditional and disaggregated space systems. From these existing methodologies, gaps in capabilities for disaggregated system design can be identified. The first major gap is that no existing methodologies simultaneously optimize for payload parameters, payload distribution, and orbital parameters. Tools like GT-FAST, VASSAR, and SODA optimize payload distribution within a cluster or to a predefined set of orbits but do not optimize payload or orbital parameters. TAT-C optimizes payload and orbital parameters but not the disaggregation scheme or payload distribution. DISCO optimizes payload and orbital parameters for a monolithic system and a predefined disaggregated architecture but does not explore the design space of levels of disaggregation. The second major gap is that existing methodologies do not include general resilience formulations. GT-FAST assesses resilience by calculating the average loss in performance for a single module failure, but this method is not extensible to systems in multiple orbits whose performance is dictated by coverage metrics. VASSAR examines coverage-based performance for single satellite loss, but this method becomes computationally prohibitive for large numbers of satellites or multiple satellite losses. These gaps in capability are addressed in the DISCO-Tech methodology. DISCO-Tech combines functionality such as the

ability to reconfigure constellations (a feature in TAT-C) and the ability to generate optimal launch manifests for disaggregated systems (a feature in DISCO). Existing applications of disaggregated systems were discussed, including multi-domain systems. The DISCO-Tech architecture is capable of assessing performance for various applications, several of which will be explored in later chapters.

Chapter 3

Methodology

DISCO-Tech uses a modular approach to allow the expansion of the code as the needs of end users change. An outline of the typical interaction between high-level modules is shown in Fig. 3.1. Decision variables, constraints, mission objectives, assets, and optimization properties are defined prior to the beginning of the optimization process during the mission definition step. Assets are typically spacecraft and ground stations, but additional asset types, such as ground or air-based sensors, can be included if suitable dynamics models are provided. During the optimizer step, a variable-length genetic algorithm is used to create candidate solutions and repair infeasible solutions. Constraints that were not repaired by the optimizer are evaluated for each solution and a penalty assigned if the constraints are violated. The propagation module determines the position of the assets over time. The mass of the satellites is estimated in the sizing module based on the payload parameters. The fitness evaluation step evaluates candidate solutions for fitness based on the designated objectives. The fitness values are passed back to the optimizer, new solutions are generated, and the process repeats until a termination criterion is met. The ordering of the modules is flexible—if the propagator depends on the satellite mass estimate (e.g. when modelling drag), the modules can be swapped.

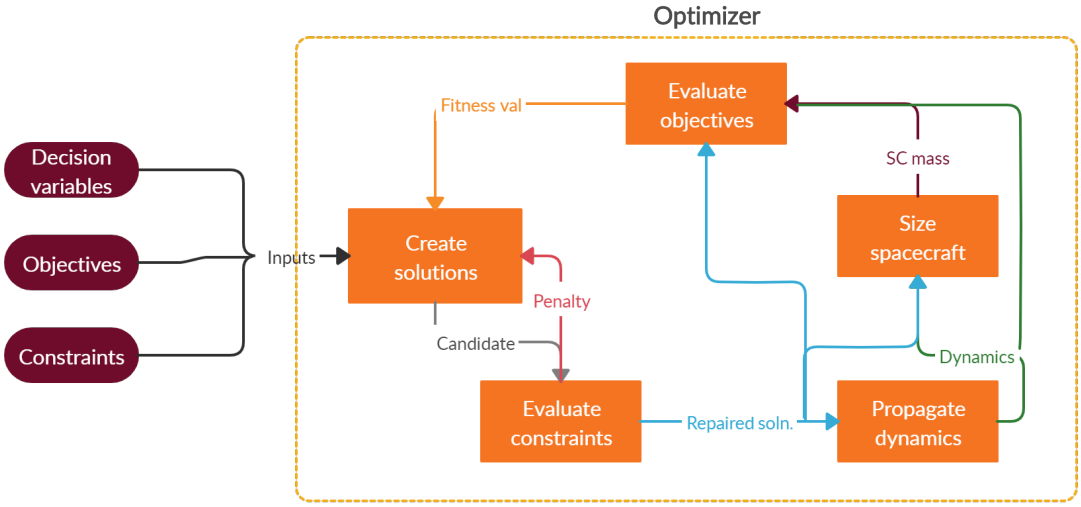


Figure 3.1: Module interaction during optimization.

3.1 Mission definition

The term "mission" in the DISCO-Tech vernacular refers to a collection of assets, objectives, constraints, and other functions that provide the information necessary to evaluate the fitness of candidate solutions. A graphical representation of a sample mission is shown in Fig. 3.2.

Assets provide a convenient way to organize spacecraft and orbital parameters. Assets are defined prior to beginning an optimization routine. Each asset, such as a satellite or ground station, contains information about the asset's path (e.g. orbit or location) and structure (physical properties and subsystem inclusions). Each asset also has a dynamics function that described that asset's motion over time. Subsystems are defined separately and may be assigned to multiple assets.

Objective functions, constraint functions, and helper functions are also parts of the mission definition. Function handles and input parameters are defined for each. Commonly used functions are defined in sections 3.5 and 3.6.

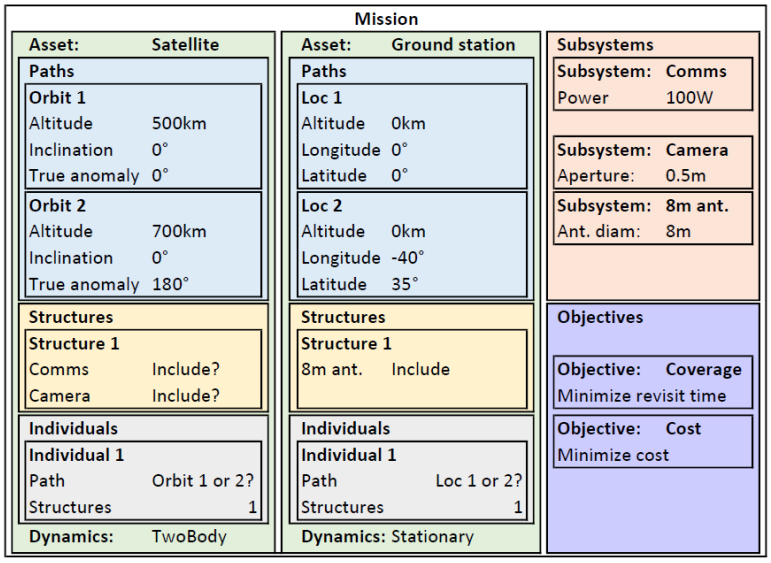


Figure 3.2: Sample mission definition.

3.2 Optimizer

3.2.1 Optimizer requirements

The DISCO-Tech methodology was developed to solve satellite and constellation design optimization problems. The optimizer must therefore be capable of handling common aspects of space mission design, such as

1. Optimization using common space mission objectives (monetary cost, coverage metrics)
2. Optimization with respect to custom user-defined objectives
3. Optimization with respect to multiple, competing objectives
4. Subsystem/payload assignment for heterogeneous space systems
5. Ability to represent varying numbers of systems

6. Constraint handling for common space mission constraints (launch vehicle capacity, link budget, etc.)

3.2.2 Problem classification

To select an optimizer, it is necessary to classify the optimization problem. The classification for a particular optimization problem will depend in part on the objective functions. It can be shown by example that coverage metrics, one of the most common objective functions for constellation design, can be not only nonsmooth but discontinuous.

The trivial constellation design problem of determining the optimal placement of a geostationary satellite such that the average percent access for a set of ground points is maximized is used for validation. The ground points of interest are on the equator at longitudes of 180°W, 60°W, and 60°E. The decision variable is the nadir longitude of the satellite. If no elevation constraints or other access constraints are imposed, the geostationary satellite can access equatorial points that are within 81 degrees of its nadir point. If no perturbations are considered, geostationary satellites have no time dependence for their access. Therefore, a ground point is either covered or it is not. For this example, the satellite will cover one or two ground points depending on its location. The resulting average percent access is shown in Fig. 3.3. It is apparent that any point for which the satellite covers two ground points maximizes coverage. Though this example is extreme, it demonstrates the potential for discontinuities in coverage metrics even in trivial cases. A similar, less trivial example could be constructed for a constellation of satellites to demonstrate that changing orbital parameters affect accesses in a discontinuous manner. Additionally, it may be difficult or impossible to analytically construct a gradient function for a given objective, requiring gradient-based optimization methods to use numerically approximated gradients. As a result, the general

DISCO-Tech problem definition must include nonsmooth, nonlinear objective functions.

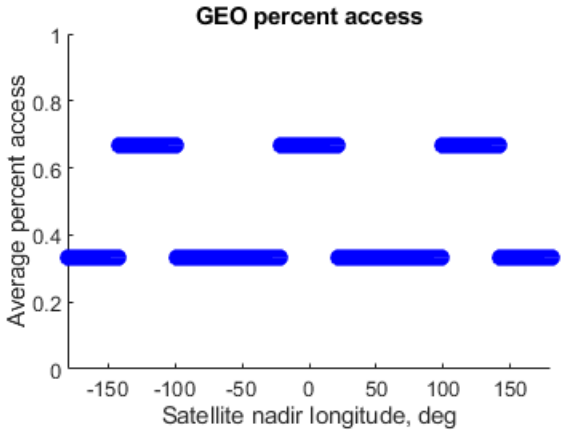


Figure 3.3: Average percent access for simple GEO sweep in nadir longitude.

In addition to the nonlinearity of the objective functions, the decision variables for satellite and constellation design are often discrete. When considering heterogeneous constellations, payload assignment is typically represented as a binary decision variable or as a continuous variable that is rounded to zero or one. Walker constellation design has the number of planes, number of satellites per plane, and spacing as integer decision variables. Other payload parameters, such as frequency and bandwidth, may be restricted to discrete values based on available equipment or government regulations.

From these considerations and the requirements outlined above, it is apparent that the framework must be capable of handling constrained multi-objective mixed variable nonlinear programming problems. Mixed variable (MV) here refers to the fact that the decision variables are a combination of continuous variables and discrete variables (including binary and integer variables). The nonlinear objective functions have no guarantees of convexity, smoothness, or continuity. The generic optimization formulation can be written as:

Minimize:

$$f_i(\mathbf{x}, \mathbf{y}), i = 1, 2, ..., n_{obj}$$

such that:

$$g_j(\mathbf{x}, \mathbf{y}) \leq 0, j = 1, 2, ..., n_{ineq}$$

$$h_k(\mathbf{x}, \mathbf{y}) = 0, k = 1, 2, ..., n_{eq}$$

$$x_m \in D_m, m = 1, 2, ..., n_d$$

$$l_p \leq y_p \leq u_p, p = 1, 2, ..., n_c$$

where:

$f_i(\mathbf{x}, \mathbf{y})$ = i-th objective function
$g_j(\mathbf{x}, \mathbf{y})$ = j-th inequality constraint function
$h_k(\mathbf{x}, \mathbf{y})$ = k-th equality constraint function
$\mathbf{x}$ = vector of discrete decision variables x_m
$\mathbf{y}$ = vector of continuous decision variables y_p
D_m = discrete set of admissible values for the m-th discrete decision variable
l_p = lower limit for the p-th continuous decision variable
u_p = upper limit for the p-th continuous decision variable
n_{obj} = number of objective functions
n_{ineq} = number of inequality constraint functions
n_{eq} = number of equality constraint functions
n_d = number of discrete decision variables
n_c = number of continuous decision variables

Many discrete design variables in constellation design are integers. Discrete, noninteger variables can always be replaced with some set of integer, binary, and continuous variables. Without loss of generality, it can be assumed that all of the discrete variables $\mathbf{x}$ are integer variables. The formulation above can instead be written as

Minimize

$$f_i(\mathbf{x}, \mathbf{y}), i = 1, 2, ..., n_{obj}$$

such that

$$g_j(\mathbf{x}, \mathbf{y}) \leq 0, j = 1, 2, ..., n_{ineq}$$

$$h_k(\mathbf{x}, \mathbf{y}) = 0, k = 1, 2, ..., n_{eq}$$

$$l_m \leq x_m \leq u_m, m = 1, 2, ..., n_d$$

$$x_m \in \mathbb{Z}, m = 1, 2, ..., n_d$$

$$l_p \leq y_p \leq u_p, p = 1, 2, ..., n_c$$

The general constellation design problem can then be classified as a multi-objective mixed integer nonlinear program (MOMINLP).

3.2.3 Optimizer selection

The suitability of common optimization techniques for solving a MOMINLP is assessed. The first aspect of the problem to be considered when selecting the main DISCO-Tech optimizer is the nonlinearity and nonsmoothness of the objective function. Numerical methods for finding minima of nonlinear functions include gradient descent methods, convexification, interior point methods, quasi-Newton methods, sequential quadratic programming, Nelder-Mead simplex, stochastic approximation, sampling methods, and metaheuristics. Because the objective function is nonsmooth and potentially discontinuous, methods that require derivative calculations such as gradient descent and sequential quadratic programming are ill-suited for this problem. Convexification, which divides the objective function into segments that can be approximated using convex estimators, may be able to divide a nonsmooth function into smooth segments that could be searched using gradient-based techniques but at the expense of increased error, introduction of infeasible solutions, or reduction in model complexity [7, 18, 147]. Alternately, gradients and derivatives can be approximated using numerical methods, though the potential presence of discontinuities poses difficulties for these methods. However, gradient-based techniques lack global exploration capabilities. Although Quasi-Newton, interior point, and Nelder-Mead solvers do not require gradient calculations,

they also lack global exploration capabilities. These methods are therefore heavily influenced by an initial guess. Multiple initial guesses can be used with these local optimization methods to explore the global solution space at the expense of increased computation time.

Simultaneous Perturbation Stochastic Approximation (SPSA) is a numerical method that approximates the gradient of a single objective function. A random perturbation vector is added to and subtracted from the current state vector, then the gradient is estimated from the objective function value at the two perturbed points. The approximate gradient is used to generate a new state [138]. This method offers probabilistic convergence guarantees to global optima for objective functions that meet certain criteria including thrice-continuous differentiability [107]. However, potential discontinuities in the objective space for constellation design remove the optimality guarantees and make gradient estimation unreliable, so SPSA is not considered further. Nonetheless, desirable features of SPSA such as constraint handling via penalty function [162], formulations with variable bounds [144], and formulations with discrete variables [163] make SPSA a good optimizer candidate for continuous objective functions that can be considered in future iterations of DISCO-Tech.

Sampling methods explore the design space by evaluating points throughout the space according to some selection methodology. One sampling method involves uniform discretizing the space of continuous variables and evaluating the objectives at the discretized points. This method expends computational resources evaluating regions of the design space that are not optimal. Additionally, optima located between the discretized points would not be found. Random sampling methods, such as Latin hypercube sampling, similarly suffer from exploration of sub-optimal space and a lack of guaranteed optimality. The DIviding RECTangles (DIRECT) search method explores the design space by identifying potential regions of optimality [81]. The potentially optimal regions are divided into smaller regions, which are searched. Division continues until convergence or until a threshold on the num-

ber of function evaluations has been reached. DIRECT is guaranteed to converge upon the global optimum for continuous functions. It can be paired with a local optimizer to improve convergence [43, 81]. Because these sampling methods do not impose convexity or continuity requirements on the objective functions, they satisfy requirements #1 and #2 for the optimizer.

Global metaheuristic techniques, including genetic algorithms [76], particle swarms [83], and simulated annealing, are capable of global exploration for nonsmooth functions through the use of non-deterministic techniques to find near-optimal solutions. Due to their global exploration capabilities and ability to handle discontinuous objective functions, a metaheuristic approach is chosen for DISCO-Tech. Since they are not limited in the types of objectives they can accept, metaheuristics permit the definition of a wide variety of user-defined objective functions. A key feature of DISCO-Tech is its modularity—the fact that the objective functions can be modified or replaced in metaheuristic algorithms without limitation on the type of function or number of inputs or outputs is a strong point in their favor. Metaheuristics therefore satisfy requirements #1 and #2 for the optimizer. Only metaheuristics and sampling methods are considered further for the general DISCO-Tech optimizer, but other solvers can be employed in post-processing to refine solutions in locally convex regions.

The next aspect of the problem to be considered is the requirement to optimize over multiple objectives. Though it is possible to transform a multi-objective optimization problem into a single objective optimization problem by performing a weighted average of the objectives, the selection of weights impacts the results. It is therefore preferable to use an optimizer that maintains an archive of multiple solutions and uses the values for each objective, rather than a weighted average. Simulated annealing produces only a single solution and is therefore not as good a fit for multi-objective optimization problems as metaheuristics that natively permit multiple solutions. Likewise, DIRECT requires a single objective function. Multi-

objective formulations exist for both genetic algorithms and particle swarm algorithms, so both options satisfy requirement #3 for the optimizer. DIRECT is not considered further due to its inability to natively handle constraints and multiple objectives, but future studies could compare the performance of DIRECT to metaheuristic techniques for constellation design.

The next aspect of the problem to be considered is the presence of integer decision variables. There are several methods for handling integer variables, including branch and bound (or similar methods), full enumeration, and directly using integer variables in a global metaheuristic algorithm. Branch and bound requires the solution of subproblems with the integer variables either fixed or allowed to take noninteger values. For constellation design variables such as number of satellites, noninteger values are nonsensical. Additionally, the optimality of branch and bound is not guaranteed with nonconvex problems. Furthermore, branch and bound is not easily extensible to multi-objective problems. As a result, branch and bound is unlikely to offer benefits over full enumeration. Full enumeration, on the other hand, is computationally expensive. Because a constellation design problem can include many continuous variables, each step of the enumeration requires solving a complex mathematical problem. The full enumeration can be intractable even with simple objective functions if the number of discrete variables is great. A constellation of ten satellites with five payload options per satellite would create over 10^{15} possibilities. By contrast, a genetic algorithm can handle integer variables through the use of crossover and mutation operations that preserve integrality. Indeed, the first GA used binary representations of decision variables [71]. The discrete variables are therefore optimized alongside the continuous variables, allowing the design space to be explored in all dimensions at once rather than sequentially through fixing the discrete variables. A GA can therefore satisfy optimization requirement #4. The code complexity is also reduced by using just a GA rather than combining a GA with enumer-

ation of the discrete variables. Discrete particle swarm optimization algorithms have also been developed, albeit more recently, including those in Refs. 101 and 27.

Genetic algorithms are more widely used for constellation design than particle swarm algorithms. Furthermore, crossover techniques exist for GAs that vary the length of the chromosome and therefore the number of variables contained in a solution [99]. This technique allows for the representation of constellations of varying numbers of satellites/orbits, satisfying requirement #5 for the optimizer. Therefore, a genetic algorithm is chosen for the optimizer.

Genetic algorithms in their basic form do not handle constraints – there is in general no guarantee that the crossover and mutation operations used by the genetic algorithm will produce feasible children from feasible parents. However, there are several methods for incorporating constraints into a GA formulation. When discussing these methods, a simple example of orbit optimization is used where both the semimajor axis a and eccentricity e of an orbit are set through the optimization. Certain combinations of a and e would cause the satellite to intersect the Earth's surface. A constraint to avoid this occurrence could be written as $a(1-e) \geq r_E + \delta$, where r_E is the Earth's radius and δ is some margin.

Common methods of constraint handling for GAs include [33]:

- Discarding candidates that violate the constraints.
  - Example: Candidates where $a(1-e) < r_E + \delta$ are removed from the pool of solutions.
- Repairing candidates that violate the constraints.
  - Example: If $a(1-e) < r_E + \delta$, modify the value of e in the genome such that $e = 1 - (r_E + \delta)/a$.

- Developing crossover and mutation operations that preserve feasibility.
  - Example: Cross over a as usual, $a_{new} = ra_1 + (1-r)a_2$. Instead of crossing over e directly, cross over the periapsis altitude values and use the result to calculate e, $e_{new} = 1 - qa_1(1-e_1)/a_{new} - (1-q)a_2(1-e_2)/a_{new}$. Similarly, replace the typical mutation operator for e with a mutation for periapsis altitude and solve for e. r and q are random values in [0,1].
- Transforming the search space such that the problem becomes unconstrained.
  - Example: Replace a and e with apoapsis altitude and periapsis altitude. If the lower limit on the new variables is equal to $r_E+\delta$, the constraint is always satisfied so the problem can be treated as unconstrained.
- Adding penalty functions to the objective functions.
  - Example: If the constraint is violated, add the penalty term $r_E + \delta - a(1-e)$ to the objective function value.

The development of operations that preserve feasibility and the transformation of the search space may not be possible for all constraints and may be complex. DISCO-Tech therefore does not employ these methods. Discarding candidates that violate constraints can make it difficult to locate initial feasible solutions and can discourage exploration when there are multiple regions of feasibility, so this method is also avoided. DISCO-Tech uses the repair and penalty function methods. In the repair method, analysts employ domain knowledge to develop functions that correct infeasible candidates. Though this method is problem-specific and requires some initial investment, it ensures the generation of feasible solutions. Its use is recommended when random solutions are likely to be infeasible or produce infinite objective values, such as in the interferometry problem described in Chapter 6. The penalty function

method is easy to implement, but the optimizer may struggle to locate feasible solutions from a randomized initial population. Various forms of penalty functions are available, including a scaled value of the constraint violation ($\sum c_i g_i(x)$ if the constraints $g_i(x) \leq 0$ are violated) or the distance from the feasible region [169]. This method is useful when developing a repair function is difficult. Performance is compared for the penalty function method and the repair method for the fractionation problem in section 4.4. Using these methods, GAs can incorporate constraints and therefore satisfy optimizer requirement #6.

The optimizer selection process is summarized in Table 3.1.

Table 3.1: Optimizer selection.

Method	**Pros**	**Cons**
Gradient descent	Guaranteed convergence to optimum for convex problems.	Relies on gradient calculations. Single objective. Unconstrained. No discrete variables.
Convexification	Allows use of gradient methods for non-convex problems.	Single objective. Can introduce error or infeasible solutions.
Quasi-Newton	Numerically approximates gradient and Hessian.	Single objective. Unconstrained. No discrete variables.
Sequential Quadratic Programming	Converges to local optimum. Constrained optimization.	Single objective. Requires differentiability of objective and constraints. No discrete variables.
Nelder-Mead	No gradient use, suitable for nonsmooth functions.	No convergence guarantees. No discrete variables. Single objective. Unconstrained.
SPSA	Global optimality for certain functions. Discrete, constrained, and bound optimization.	Gradient approximation unreliable for discontinuous functions. Single objective.

Interior point	Guaranteed convergence to optimum for convex problem. Constrained optimization.	No discrete variables. Single objective.
Uniform/ random sampling	No gradient use, suitable for nonsmooth functions and multiple objectives.	Computationally expensive. No optimality guarantees.
DIRECT	No gradient use, global optimality for continuous functions.	Single objective. Unconstrained except for bounds.
Simulated annealing	Global exploration. Discrete optimization. Nonsmooth functions.	Single objective. Unconstrained except for bounds. No optimality guarantees.
Particle swarm	Multiple objectives. Nonsmooth functions.	Unconstrained except for bounds. Less heritage than GA. No variable-length methods. No optimality guarantees.
Branch and bound	Guaranteed optimality for mixed integer convex problems.	Cannot make discrete constellation variables continuous. Requires method of solving relaxed problem. Nonconvexity removes optimality guarantees.
Full enumeration	Guaranteed optimality for discrete problems.	Computationally expensive. Requires discretization of continuous variables or separate, simultaneous optimization of continuous variables.
Genetic algorithm	Multiple objectives. Nonsmooth functions. Variable-length versions. Constellation design heritage. Formulations using linear constraints.	No natural inclusion of nonlinear constraints. No optimality guarantees.

3.2.4 Optimizer description

The previous section outlines the benefits of using a genetic algorithm. This section provides information about genetic algorithms in general. It also discusses a few desirable aspects of the specific genetic algorithm used, including a multi-objective formulation, variable-length crossover operations, and constraint handling.

A genetic algorithm mimics the processes of natural selection and evolution. Each decision variables corresponds to a *gene* that can take some set of values. Each candidate solution is represented by a set of genes called a *genome* or a *chromosome*. A standard GA has the following steps:

1. **Initial population:** A set of genomes are randomly generated to form an initial *population* of solutions. This initial population makes up the first generation.

2. **Evaluation:** Each of the candidate solutions are evaluated for fitness with respect to one or more objective functions.

3. **Sorting:** The candidate solutions in the current generation are compared and ranked from best to worst, though the method of ranking differs between GAs. Candidates from previous generations may also be included in the sorting process, depending on the algorithm.

4. **Selection:** Groups of solutions are selected to act as *parents* for the next generation of solutions. The selection process typically is random but gives higher selection probabilities to better solutions.

5. **Crossover:** The candidates chosen in the previous step are combined to produce new solutions, called *children*.

6. **Mutation:** Genes in the children are randomly modified to introduce diversity into the solutions and encourage exploration of the design space.

7. The children form the next generation of solutions. Repeat steps 2–6 until convergence is achieved.

The implementation of each of these steps is discussed in detail below. Due to the particular needs of DISCO-Tech, a custom GA is developed based off of the ϵNSGA-II and Borg GAs [68, 91]. ϵNSGA-II provides the concept of epsilon boxing, the sorting operations, and the mutation operation. Borg provides the concept of an archive of solutions, the process for selecting parent solutions, a method for re-invigoration, dynamic population sizing, and the convergence criteria for the optimization. A custom crossover operation is introduced to vary the length of the genome. In constellation design, this variable-length crossover operation allows constellations of various sizes and configurations to be represented without resorting to turning segments of the genome "off" using flagging. Because the algorithm generates nondominated fronts for problems with multiple objectives, a variety of feasible solutions can be found without assigning arbitrary weights to the different objectives.

Sorting

Sets of solutions can be grouped into *nondominated fronts*, groups of solutions in which no solution is strictly better than any other solution [41]. An example group of solutions for a two-objective minimization problem is shown sorted into fronts in Fig. 3.4. The front to which a solution belongs determines its *rank*, where ranks of a lower number denote better solutions. The set of nondominated solutions that dominate all other solutions is called the *Pareto-optimal front*. For problems with continuous decision variables exhibiting trade-offs between objectives, it is possible to have an infinite number of solutions along the Pareto-

optimal front. The concept of epsilon boxing limits the number of nondominated solutions by treating solutions within the same range of objective values, called the epsilon box, as equivalent [91]. The solution closest to the dominant corner of the box is considered to be the best solution in the box. The same set of solutions is shown in Fig. 3.5 with the epsilon-nondominated front marked. The first objective has an epsilon value of 1, while the second objective has an epsilon value of 0.5.

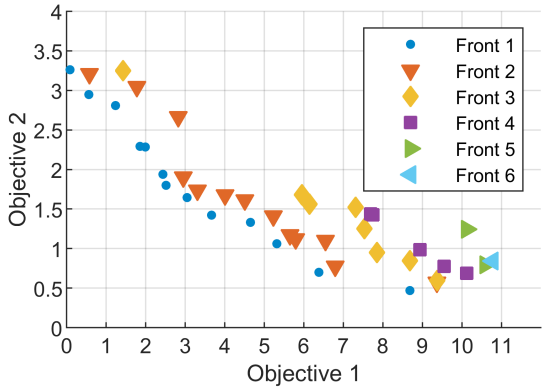


Figure 3.4: Solutions sorted into nondominated fronts for a two-objective minimization problem.

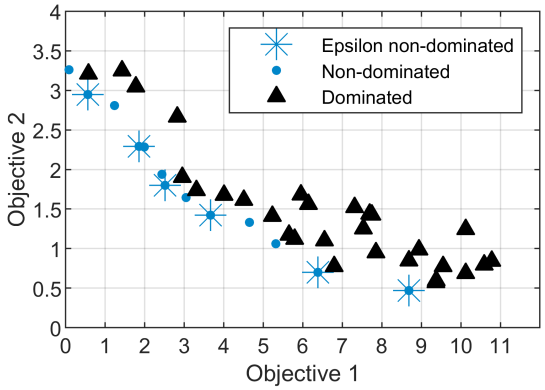


Figure 3.5: Epsilon-nondominated front for a two-objective minimization problem.

The algorithm separately maintains two sets of solutions: a population and an archive [91]. The population evolves over time based on the performance of newly generated solutions but does not contain only nondominated solutions. The inclusion of dominated solutions in the population promotes diversity. If a newly generated solution dominates one or more solutions in the current population, exactly one of those solutions is removed and replaced with the new solution. If the new solution is nondominated with respect to all solutions in the population, it is added and a random population member is removed. The population is maintained at a constant size unless a restart is triggered, as described below. Population members within the same nondominated front are then sorted by their *crowding distance*. Crowding distance is the relative proximity to other solutions, as detailed in Ref. 41.

The second set of solutions is the archive, the set of epsilon-nondominated solutions found throughout the evaluation. Only a single point per epsilon box, that which is closest to the box's dominant corner, is maintained. If a new solution is epsilon-nondominant with the archival solutions, it is added to the archive. If the new solution is epsilon-dominant with respect to one or more archival solutions, those solutions are removed from the archive and the new solution added [91].

Selection

Once the solutions have been sorted, pairs of solutions are selected to act as parents to the next set of solutions. The first parent in a pair is randomly selected from the archive. The second parent in a pair is chosen from the population through tournament selection [41]. In tournament selection, the current candidate is compared to a new candidate each round. If the candidates are of different nondominated fronts, the less dominated candidate is chosen. If candidates are of the same front, the candidate with the larger crowding distance is chosen. The number of pairs selected is equal to half the population size.

Crossover

The crossover algorithm used in this tool is a departure from the standard crossover technique used in most genetic algorithms, including ϵNSGA-II and Borg. The algorithm is designed to permit length change in the genome. A previous example of a variable length genetic algorithm is the exG algorithm developed at the California Institute of Technology [99]. Unlike exG, the crossover algorithm used by DISCO-Tech divides the genome into segments of different types and ensures that crossover will only occur between like segments. Ryerkeck refers to this genomic division in which segments of the genome represent different systems

as *metametric* [132]. The segments themselves are called *metavariables*, while variables not contained within the segments are called *global variables*. For a constellation design problem, potential metavariable types include subsystems, hardware/structures, orbits/paths, and individual satellites. Global variables for a constellation design problem can include values shared by multiple satellites, such as a common semimajor axis.

Previous work by Hitomi et. al. using variable-length genetic algorithms for constellation design did not use multiple types of metavariables and consider payload assignment. This study therefore lacked the flexibility to vary payload distribution and subsystem variety, a key feature for disaggregated system optimization [73]. DISCO-Tech can use multiple types of metavariables to simultaneously vary the payload variety, orbit variety, and satellite variety. Alternately, a single metavariable can contain information on both the orbit and payload assignments for a satellite. Through this process, space systems can be modelled without premature decisions about or bias toward particular architectures, such as monolithic systems or predefined combinations of payloads.

First, a random value is compared to the length change rate (LCR) to determine if length changing is permitted. If a length change can occur, the starting and ending segment type are chosen randomly. The segment numbers are then chosen randomly for both parents. The starting and ending position within a segment is also chosen randomly but must be the same for both parents. Furthermore, limits are imposed on the maximum and minimum number of segments of each type and selections changed if the crossover will violate these limits. Once the portions to be swapped have been determined, a weighted swap is performed between the parents, resulting in two children. The crossover operation is performed for each pair chosen in the selection step, forming the next generation.

Finally, a random value for each child is compared to the expansion rate (ER) to determine if a random segment should be added. This step prevents the optimizer from becoming

stuck with an archive of minimum length solutions. If the random value is less than or equal to the ER and there are some segments that have not reached their maximum number of segments, expansion is performed. A segment type is randomly selected, and a new segment of that type is randomly generated and added to the genome. Pseudocode for the crossover operation is shown below:

1: l =rand(1)
2: Choose starting and stopping segment types
3: i_1, i_2 =empty - indices to be swapped
4: **if** Starting and stopping types are the same **then**
5: Choose the starting and stopping segment numbers for each
6: **if** $l > LCR$ **then**
7: No length change - shift the stopping segment number of the longer segment back so it matches the length of the shorter segment
8: **else if** the swap will violate the segment number limits **then**
9: Reduce the size of the larger segment being moved
10: **end if**
11: Choose the starting and stopping position in the segment.
12: Add indices from starting and stopping positions to the first position in i_1 and i_2
13: **else**
14: Choose the starting segment numbers for the starting segment type. The stopping segment number will be the last segment of this type.
15: **if** $l > LCR$ **then**
16: No length change - shift the starting segment number of the longer segment forward so it matches the length of the shorter segment
17: **else if** the swap will violate the segment number limits **then**
18: Reduce the size of the larger segment being moved

end if

Choose the starting position in the segment.

Add indices from starting positions to the end of the segment type to the first positions of i_1 and i_2

for $j = 1$: number of segment types between starting and stopping **do**

Add indices from the start of the type to the end of the shorter type to the j+1 positions of i_1 and i_2

end for

Choose the stopping segment numbers for the stopping segment type. The starting segment number will be the first segment of this type.

if $l > LCR$ **then**

No length change - shift the stopping segment number of the longer segment backward so it matches the length of the shorter segment

else if the swap will violate the segment number limits **then**

Reduce the size of the larger segment being moved

end if

Choose the stopping position in the segment.

Add indices from the start of the segment type to the stopping positions to the last positions of i_1 and i_2

end if

r =rand(1)

child1 = parent1; child2 = parent2

for k =1: number of segment types from starting to stopping **do**

s_1 = portion of first parent corresponding to $i_1(k)$

s_2 = portion of second parent corresponding to $i_2(k)$

Extend s_1 or s_2 so lengths are equal

$s_2 = s_2 - r(s_2 - s_1)$

$s_1 = s_1 - r(s_1 - s_2)$

Shorten s_1 or s_2 back to original lengths

Remove s_1 from child1 and insert s_2 in its place

Remove s_2 from child2 and insert s_1 in its place

end for

for $c \in$ {child1, child2} **do**

r =rand(1)

if $r \leq ER$ **then**

if there are segments in c that have not reached their maximum number **then**

Randomly select a viable segment type

Generate a new segment of that type

Append the new segment to the block of segments of the appropriate type in c

end if

end if

end for

Mutation

A predefined mutation rate determines the likelihood that an individual element of the genome will mutate. For each gene in a particular solution's genome, a random number is chosen and compared to the mutation rate. If the random number is less than or equal to the mutation rate, that element of the genome is changed to a random value in the set of admissible values for that element [71]. This process is performed for every element of every genome in the new generation.

Constraint Handling

As discussed in subsection 3.2.3, DISCO-Tech handles constraints in one of two ways. The first option evaluates constraints for each genome immediately after the mutation phase. If a genome violates a constraint, it is repaired. A method must be defined a priori for correcting the violation, which is used to modify the genome to satisfy the constraint. Using this method, constraints are handled one at a time and must not introduce violations in other constraints through their repair.

The second option incorporates constraints into the evaluation function. Constraints are checked as close to the beginning of the evaluation function as possible. For each violated constraint, a penalty value is increased based on the severity of the violation. The fitness is set to the value of the penalty and the remaining code skipped. One advantage of this method is that it considers the extent to which the constraints are violated. The genetic algorithm will minimize the constraint violation, eventually pushing the solution into the admissible space. Furthermore, if no valid solutions exist, solutions will be found that minimize the constraint violations. A disadvantage of this method is that it could take a long time if no initial solutions are near the admissible space. Although this method does not currently consider the performance of infeasible solutions with respect to the objective function, it could easily be adapted to consider the inadmissible solution performance, as in Ref. 169.

Convergence

To ensure that the genetic algorithm does not run indefinitely, convergence criteria must be defined. DISCO-Tech has the ability to terminate after a certain amount of elapsed time or after a certain number of function evaluations. However, neither of these criteria assess whether the optimizer has converged upon a solution. The preferred method for stopping

evaluation is therefore the convergence criteria used in the Borg algorithm [68]. The user dictates a threshold for the maximum number of generations the algorithm is allowed to run without improvement. Improvement here is defined as the discovery of a solution that is either in a new epsilon box that is epsilon-nondominant with the rest of the archive or is in a new epsilon box that epsilon-dominates one or more solutions in the archive. Finding a better solution within an existing epsilon box is not considered an improvement for convergence purposes. Once the threshold for the maximum number of generations is reached, a restart is triggered [68].

When a restart is triggered, the population is emptied. The solutions in the archive are counted. The desired population size is the minimum population size defined by the user or four times the archive size, whichever is greater. The archive is added to the population. The remainder of the population is filled with randomly generated solutions [68]. This process injects diversity into the algorithm and provides the potential to find a better global solution if the previous run converged to a local optimum. The algorithm runs until some predefined number of restarts has been reached, after which the algorithm is considered to have converged.

3.2.5 Alternate optimizers

Because DISCO-Tech is modular, it is possible to replace the genetic algorithm with another optimizer. Other optimizers may offer superior performance for certain classes of problems, such as convex problems. Additionally, local optimizers can be used to improve local performance once the GA has explored the global design space. The optimizers that currently are built into the DISCO-Tech architecture include the Nelder-Mead simplex method [121], simulated annealing [79, 85], hill climb methods, and interior point methods [60]. These

methods do not permit variable-length representations. Constraints are handled through penalty functions or repair functions internal to the function evaluation step.

3.3 Propagation

There are two methods available for propagating satellite position over time. The first, more flexible method is a numerical propagator. Conservative forces are determined by taking the gradient of the potential. The gravitational potential energy per unit mass due to a nonspherical body can be characterized using a spherical harmonics representation [12],

$$V = -\frac{\mu}{r}\left(1 + \sum_{l=2}^{\infty}\left(\frac{r_E}{r}\right)^l \sum_{m=0}^{l}\left[\overline{C_{lm}}\cos(m\lambda) + \overline{S_{lm}}\sin(m\lambda)\right]\overline{P_{lm}}\left(\sin(\phi)\right)\right) \tag{3.1}$$

where V is the gravitational potential energy per unit mass; μ is the standard gravitational parameter for the planet; r, λ, and ϕ are the radius from the planet's center to the satellite, the longitude of the satellite, and the geocentric latitude of the satellite; r_E is the reference radius of the planet used to develop the model; $\overline{C_{lm}}$ and $\overline{S_{lm}}$ are coefficients; and $\overline{P_{lm}}(\sin(\phi))$ is the normalized associated Legendre polynomial of degree l and order m, evaluated for $\sin(\phi)$. The propagator currently truncates the series after the fourth degree terms ($l = 4$) to avoid excessive computation time, though additional terms can be included as needed. The magnetic field of a body could be modeled in a similar manner. The spherical gradient of the potential is taken to obtain the conservative force in the spherical coordinate frame. The force can then be rotated to the planet-centered inertial frame prior to integration. Nonconservative forces, such as drag and control forces, can also be modeled and rotated to the inertial frame. The equations of motion are then integrated numerically using a three-

stage, third-order, Runge-Kutta method as the ordinary differential equation (ODE) solver. By controlling the tolerances of the ODE solver and the force and potential models, satellite positions can be solved with a high degree of accuracy. Furthermore, numerical propagators can be developed for other vehicle types and the equations of motion solved simultaneously. However, this method is computationally expensive and becomes infeasible for large numbers of satellites.

The second method uses the Lagrange variation of parameters method to determine the secular effects of the low-degree zonal harmonic terms on the osculating orbital elements over time [108, 154]. The secular effect on the semimajor axis a, eccentricity e, and inclination i are negligible, so those parameters will be treated as constant in time. The rate of changes of the longitude of the ascending node, Ω, the argument of periapsis, ω, and the mean anomaly, M, due to the spherical planet model and second and fourth order terms are

$$\begin{aligned}\dot{\Omega}_{sec} = &-\frac{3J_2 r_E^2 n\cos(i)}{2p^2} \\ &+\frac{3J_2^2 r_E^4 n\cos(i)}{32p^4}\left(12-4e^2-(80+5e^2)\sin^2(i)\right) \\ &+\frac{15J_4 r_E^4 n\cos(i)}{32p^4}\left(8+12e^2-(14+21e^2)\sin^2(i)\right)\end{aligned} \tag{3.2}$$

$$\begin{aligned}\dot{\omega}_{sec} = &\frac{3nJ_2 r_E^2}{4p^2}\left(4-5\sin^2(i)\right) \\ &+\frac{9nJ_2^2 r_E^4}{384p^4}\left(56e^2+(760-36e^2)\sin^2(i)-(890+45e^2)\sin^4(i)\right) \\ &-\frac{15J_4 r_E^4 n}{128p^4}\left(64+72e^2-(248+252e^2)\sin^2(i)+(196+189e^2)\sin^4(i)\right)\end{aligned} \tag{3.3}$$

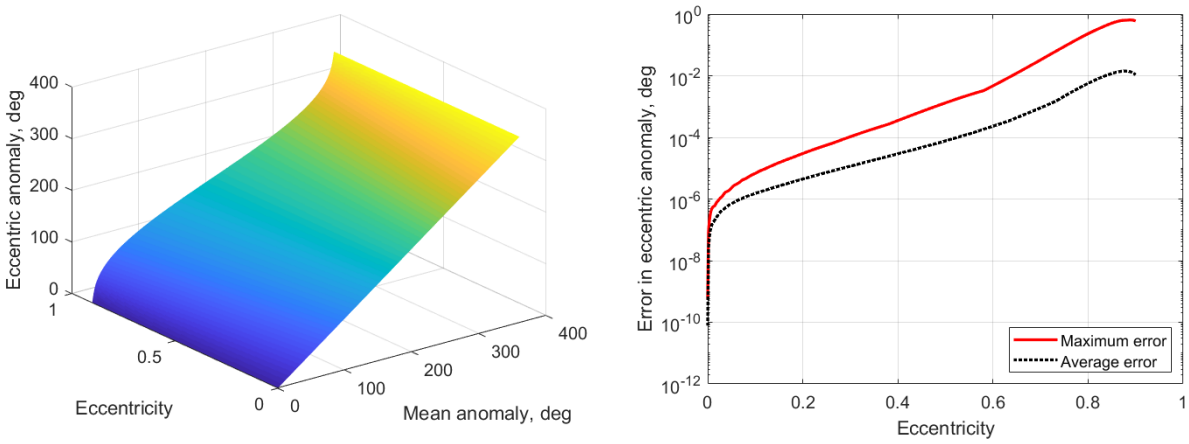
$$\dot{M} = n+\frac{3nr_E^2 J_2\sqrt{1-e^2}}{4p^2}\left(2-3\sin^2(i)\right) \tag{3.4}$$

$$+\frac{3nr_E^4J_2^2}{512p^4\sqrt{1-e^2}}\Big(320e^2-280e^4+(1,600-1,568e^2+328e^4)\sin^2(i)$$
$$+(-2,096+1,072e^2+79e^4)\sin^4(i)\Big)$$
$$-\frac{45J_4r_E^4e^2n\sqrt{1-e^2}}{128p^4}\Big(-8+40\sin(i)-35\sin^2(i)\Big)$$

where $p=\frac{h^2}{\mu}=a(1-e^2)$ is the semi-parameter, n is the mean motion, and the J_2 and J_4 constants are the negatives of the second and fourth degree non-normalized zonal harmonic coefficients. The rates are constant, so the values of Ω, ω, and M can be found by multiplying the rates by the time t and adding the result to the initial value. It is then necessary to find the eccentric and true anomalies, E and ν, to calculate the Cartesian spacecraft position and velocity. The relationship between the mean and eccentric anomalies is described by Kepler's equation, $M=E-e\sin(E)$. Since Kepler's equation is transcendental in E, the equation must be solved numerically. Because solving Kepler's equation at every time step for every satellite would be cumbersome, the solution is precomputed for a 50x50 grid of mean anomalies and eccentricities. Bicubic interpolation is then used to find the eccentric anomaly for any combination of mean anomaly and eccentricity. The interpolation surface and its error relative to the actual solution to Kepler's equation are shown in Fig. 3.6. The true anomaly can then be computed from the eccentric anomaly.

3.3.1 Maneuvering

DISCO-Tech supports impulsive maneuvering and allows the optimizer to set the magnitude and direction of the maneuver. Propagation occurs until the time of a maneuver is reached. The changes in velocity Δv imparted by the maneuvers are then added to the velocities of the corresponding satellites. The position remains unchanged, so the old position and new velocity are used as initial conditions for further propagation. Alternately, changes in



(a) Bicubic interpolation surface.

(b) Error in bicubic interpolation.

Figure 3.6: Bicubic interpolation of Kepler's equation.

orbital elements can be provided, though this method of definition can introduce discontinuities in position. Propagation then continues until the next maneuver time is reached. It is also possible to model more complex control schemes if they are of a well-known form or are analytically derived. Although a large body of research exists with regards to optimal maneuvering of spacecraft, the inclusion of an optimization routine for each satellite's maneuvers is generally computationally prohibitive due to the nesting of optimization problems. Subsection 3.5.5 discusses one such control scheme that is solved via linear programming, though the results are approximate. If the maneuvering duration is sufficiently short, the linear program may be solved without introducing large computation time into the function evaluation.

3.4 Sizing

Payload and spacecraft mass are the main drivers of mission cost, one of the most common objectives in space system design. Mass estimation early in the design process approximates

sizing information based on design variables such as payload power, antenna size, and orbit. This process relies heavily on analogies with existing and historic systems. Each subsystem is sized in the DISCO-Tech architecture—the inputs to each sizing module can be optimization decision variables, user inputs, or values derived earlier in the DISCO-Tech workflow.

3.4.1 Payload: Communications

The communications subsystem as defined by DISCO-Tech includes the hardware for telemetry, tracking, and command (TT&C); data downlinking; and communications relay. Fig. 3.7 shows the inputs and outputs for the module. Generally, the link margin and rain rate are constants. The antenna dimensions may be constants or decision variables. The data per orbit is either a constant (for a communications satellite with a desired data throughput) or a function of other payload variables that change the amount of data generated.

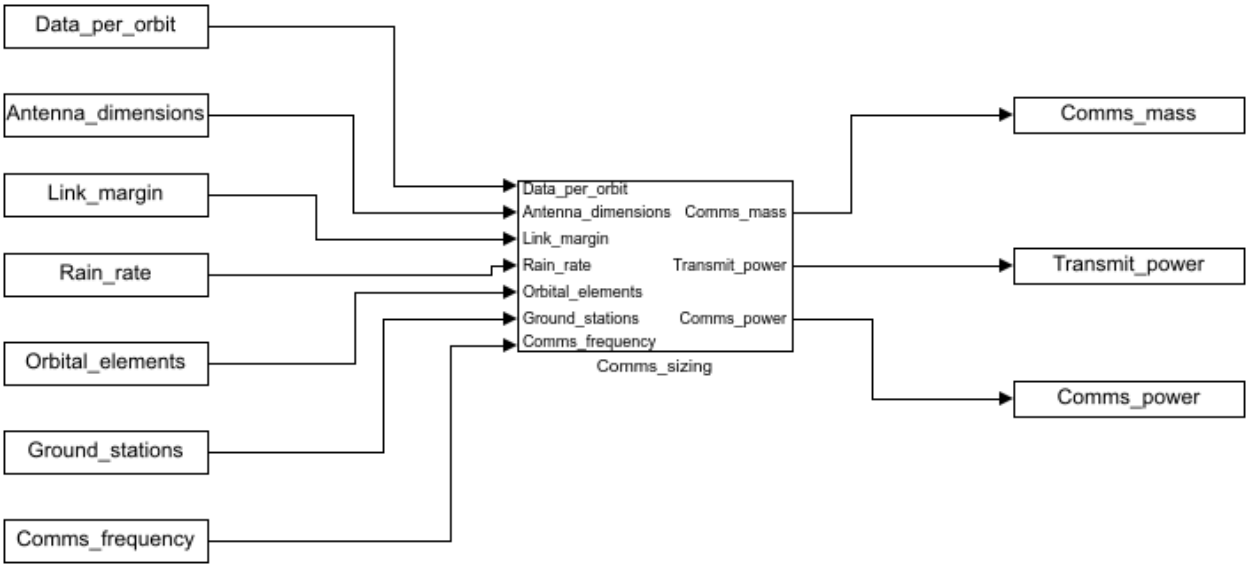


Figure 3.7: Block diagram for communications sizing module.

The process for sizing the subsystem is:

1. **For data downlinking only:** Given a list of ground station positions and receiver

gains, calculate the average percent access per ground station per orbit according to the process described in subsection 3.6.1. Assuming that accesses for different ground stations do not overlap, the total percent access per orbit is equal to the average percent access times the number of ground stations. The data generation rate is equal to the data per orbit divided by the orbital period. The required communications subsystem data rate is therefore equal to the data generation rate divided by the total percent access per orbit.

2. **For communications satellites:** The required data rate is a function of the number of users per satellite and the number of satellites in the system. The required data rate per user is supplied as an input. For a constellation with a known distribution of users, the maximum number of simultaneous users per area is also provided. The area within the satellite field of view is calculated and is used to determine the number of users simultaneously served by the satellite. If the field of view overlaps with that of other satellites, the users are divided between the two satellites. If the distribution of users is unknown, it is assumed that the users are evenly distributed across all satellites. Furthermore, when no additional data is available, it is assumed that communication between users anywhere in the covered region is equally likely. Assuming that satellites can crosslink with satellites one ahead or behind them within their orbital plane and with the closest satellite in immediately adjacent planes, the average number of links required to support a connection is calculated. The total simultaneous data rate is estimated as the product of the number of simultaneous users per satellite, the data rate per user, and the average number of links required for a connection. The crosslinking system is assumed to use the same frequency as the overall communications subsystem and does not have its hardware modelled separately.

3. The satellite is assumed to use a quadrature phase-shift keying (QPSK) modulation

scheme with a roll-off factor of $\alpha = 0.2$ and a code rate of $\rho = 0.9$, though these parameters can be modified as needed. The spectral efficiency e_s is

$$e_s = m\rho/(1+\alpha) \tag{3.5}$$

Since QPSK uses two bits per symbol ($m = 2$), the spectral efficiency for the default modulation scheme is 1.5 bits per second per Hertz. The required bandwidth for the system is equal to the communications data rate divided by the spectral efficiency. For multiple access scenarios such as voice communications, a frequency division multiple access (FDMA) scheme is employed and the required bandwidth is increased by an additional 10% to account for the necessary guard bands.

4. The data rate R and bandwidth BW values can be used to calculate the necessary energy per bit to noise power spectral density ratio (E_b/N_0). The theoretical limit on the maximum spectral efficiency for a given E_b/N_0 in decibels is given by the Shannon-Hartley theorem,

$$R \leq BW \log_2 \left(1 + 10^{E_b/N_0/10} \frac{R}{BW}\right) \tag{3.6}$$

However, this limit is not achievable in practice. Instead, DISCO-Tech estimates the required E_b/N_0 in one of two ways. If the desired bit error rate (BER) is known, it can be used to calculate the required E_b/N_0 for a particular coding scheme. A Reed-Solomon code with a message length of 223 bits and a block length of 255 bits is used as the encoding scheme. Matlab's Communications toolbox is used to calculate the needed E_b/N_0 for the desired BER and coding scheme. If the desired BER is not provided, DISCO-Tech uses the E_b/N_0 required for link closure as dictated by the Digital Video Broadcasting-Satellite, Second Generation (DVB-S2) standard, plus a link margin as defined by the user. For the default modulation scheme, that threshold

value is 3.89 dB [165].

5. The link budget for the satellite is then formed for the most distant serviced point. That point is dictated either by the minimum allowable elevation angle for access or by the satellite field of view (as dictated by the antenna half-power beamwidth) if the satellite does not point at the target. The link budget is satisfied if the calculated E_b/N_0 is greater than or equal to the required value. E_b/N_0 in decibels is calculated as

$$\frac{E_b}{N_0} = 10\log_{10}(P_t) + G_t + G_r + L_l + L_s + L_a + L_i + L_p - 10\log_{10}(kT_sR) \tag{3.7}$$

where

P_t = transmit power per channel, W
G_t = transmit antenna gain, dBi
G_r = receive antenna gain, dBi
L_l = total line loss and backoff, dB
L_s = space loss, dB
L_a = atmospheric loss, dB
L_i = implementation loss, dB
L_p = pointing loss, dB
k = Boltzmann's constant, J/K
T_s = system noise temperature, K
R = data rate per channel, bps

The antenna gains are calculated based on their aperture areas. For a parabolic antenna with an efficiency of 0.55, the peak gain is $G_t = 17.8 + 10\log_{10}(D_iD_c) + 20\log_{10} f$, where D_i and D_c are the in-track and cross-track antenna diameters in meters and f is the frequency in GHz. The resulting gain is in dBi.

Typical loss values are taken from Space Mission Engineering: The New SMAD [165]. The space loss in dB is

$$L_s = 147.55 - 20\log_{10}(d) - 20\log_{10}(f \times 10^9) \tag{3.8}$$

where d is the path length in meters.

The normalized electric field strength of a parabolic antenna with diameter D and a uniformly illuminated aperture is [93]

$$E(\theta) = \frac{2\lambda}{\pi D} \frac{J_1[(\pi D/\lambda)\sin(\theta)]}{\sin(\theta)} \tag{3.9}$$

where θ is the angle the vector between the satellite and target makes with the antenna boresight vector. The power per unit area S is proportional to the square of the electric field strength. Since the power per unit area is equal to $P_t G_t/(4\pi d^2)$, the gain is therefore also proportional to the square of the electric field strength. The difference between the peak antenna gain and the antenna gain at an angle θ off-boresight is therefore equal to $L_p = E(\theta)^2$. If the antenna/satellite aims at a ground target, θ is equal to the satellite's pointing error as dictated by the attitude control system. If instead the antenna is nadir pointing, θ is equal to half the half-power beamwidth of the satellite or to an effective antenna field of view as dictated by an elevation angle constraint on the ground station, whichever is smaller.

The attenuation for a transmission in the nadir direction is calculated from the estimates provided by ITU-R Recommendation P.676-9 [2]. The distance the signal travels through the atmosphere is calculated using trigonometry, as signals travelling off-nadir will travel farther in the atmosphere. Assuming the atmosphere is 16 km thick, the ratio between the distance actually travelled through the atmosphere and the nadir distance is found and multiplied by the nadir attenuation to get the total attenuation due to gases in the atmosphere. Additionally, attenuation due to rain is calculated using the built-in Matlab function *rainpl* with a user-supplied rain rate.

Finally, the system noise temperature is estimated using receiver noise figures and line

loss temperatures from reference systems and antenna noise temperatures due to rain noise and sky noise as given in Refs. 98 and 165.

Given input system parameters and a desired E_b/N_0 level, Eq. 5 is solved to obtain the required transmit power per channel.

6. For communications satellites, 36 MHz transponders are assumed. The number of transponders is therefore equal to the system bandwidth divided by 36 MHz. For data downlinking systems, the number of transponders is assumed to be two. The system is assumed to use Travelling Wave Tube Amplifiers (TWTA). The number of TWTAs is at least as great as the number of transponders. The TWTAs are assumed to provide at most 70 W of power, potentially increasing the number of TWTAs beyond the number of transponders. TWTA mass and efficiency are calculated based on the needed power using the curves in Ref. 98. The transponders are assumed to have a mass of 2.875 kg [165]. The transmitter mass is calculated as the sum of the transponder and TWTA masses, plus multiplexer masses of 1.16 kg per transponder. The connectors are assumed to increase the mass by 40%.

7. The antenna mass is estimated based on analogies with existing systems as $m_{ant} = 6\text{kg/m}^2 D_i D_c$ for a parabolic antenna [98].

8. The total subsystem mass is therefore equal to the sum of the transmitter, antenna, and connector masses. The transmit power is as calculated by the link budget. The peak power used by the communications subsystem is equal to the transmit power divided by the TWTA efficiency.

3.4.2 Payload: Radar

Historical data for satellites with traditional radar and SAR payloads are used to approximate the mass for a radar payload. These data are shown in Table 3.2[1]. Analysis of the data showed that the peak payload power was a chief driver of subsystem mass. Additionally, SAR systems with a scanning (scanSAR) mode are more complex and therefore more massive. The data set was split between radars with a scanning mode and radars without a scanning mode.

Table 3.2: Radar mission data.

Mission	Peak power, W	In-track dim, m	Cross-track dim, m	Ant. area, m^2	Freq., GHz	SAR?	Scan?	Payload mass, kg
Terra SAR-X	2240	4.80	0.70	3.36	9.65	True	True	400.00
ERS-1/2	4800	10.00	1.00	10.00	5.30	True	False	326.00
Envisat	1365	10.00	1.30	13.00	5.33	True	True	832.00
Sentinel-1A-D	4400	12.30	0.82	10.10	5.41	True	True	945.00
Meteor-M N1/2	1000	13.40	0.25	3.35	9.62	True	False	150.00
Seasat	1000	10.70	2.10	22.47	1.28	True	False	147.00
Magellan	325	3.70	3.70	10.75	2.39	True	False	154.00
KOMP SAT-5	1700	4.48	0.70	3.14	9.66	True	True	520.00
Biomass	360	12.00	12.00	113.10	0.44	True	False	202.00
ALOS	2000	8.90	3.10	27.59	1.27	True	True	600.00
ALOS-2	6100	9.90	2.90	28.71	1.26	True	True	656.80
RISAT-1/1A	2880	6.29	2.09	13.15	5.35	True	True	950.00
COSMO SkyMed	1250	5.70	1.40	7.98	9.60	True	True	650.00
RADARSAT CM	1270	6.88	1.37	9.43	5.41	True	True	400.00
Jason-3	33	1.20	1.20	1.13	13.58	False	False	70.00
SeaSat	2000	1.00	1.00	0.79	13.56	False	False	93.80
SARAL	2	1.00	1.00	0.79	35.75	False	False	40.00

[1]Data compiled from EOPortal.org, NASA Space Science Data Coordinated Archive, and the WMO Observing Systems Capability Analysis and Review Tool.

To explain the remaining variations in mass, various parameters were examined to determine their influence on the mass. The antenna area and frequency were found to have the greatest impact. Antenna area directly affects the mass of the antenna. Frequency's contribution is less obvious but may be a result of the availability of parts for high-frequency radar applications, increased bandwidth or data rate at high frequencies, or increased beam precision needed due to narrower beams from the same antenna area. As a result, the product of antenna area and frequency is used as an additional driver. The payload mass is plotted as a function of the described parameters in Fig. 3.8.

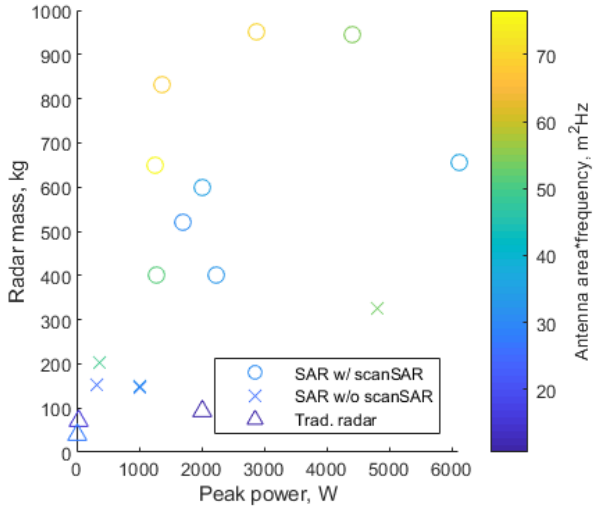
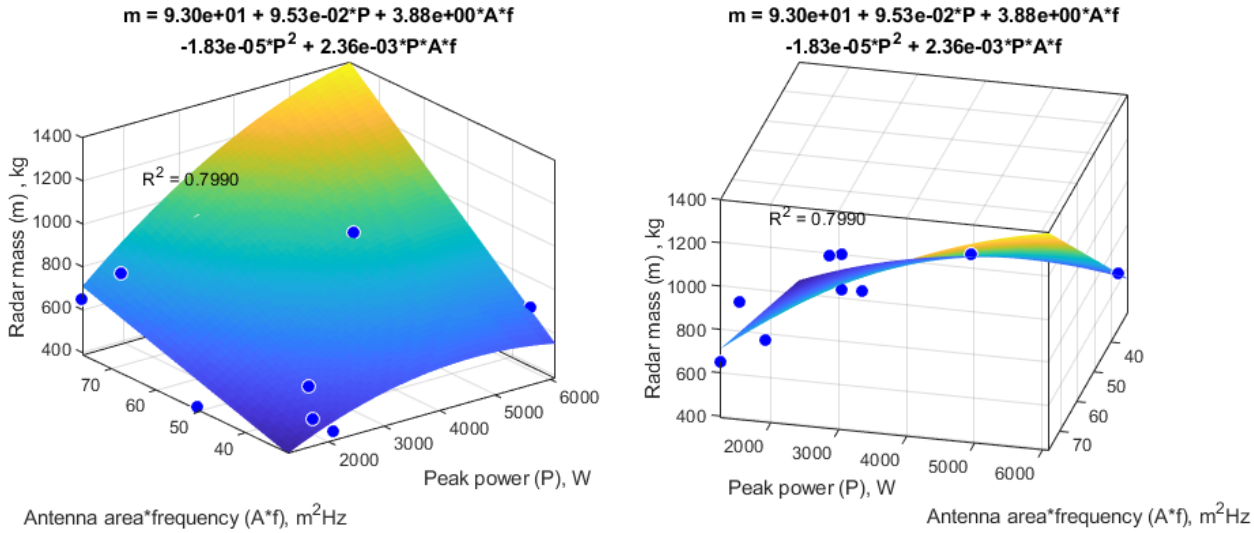


Figure 3.8: Radar mass as a function of system parameters.

A polynomial surface was fit to the scanning data set. The fit had a degree of two in the x dimension (peak power) and a degree of one in the y dimension (antenna area times frequency). The fit has an R^2 value of 0.799 and is depicted from two different points of view with the relevant data set in Fig. 3.9.

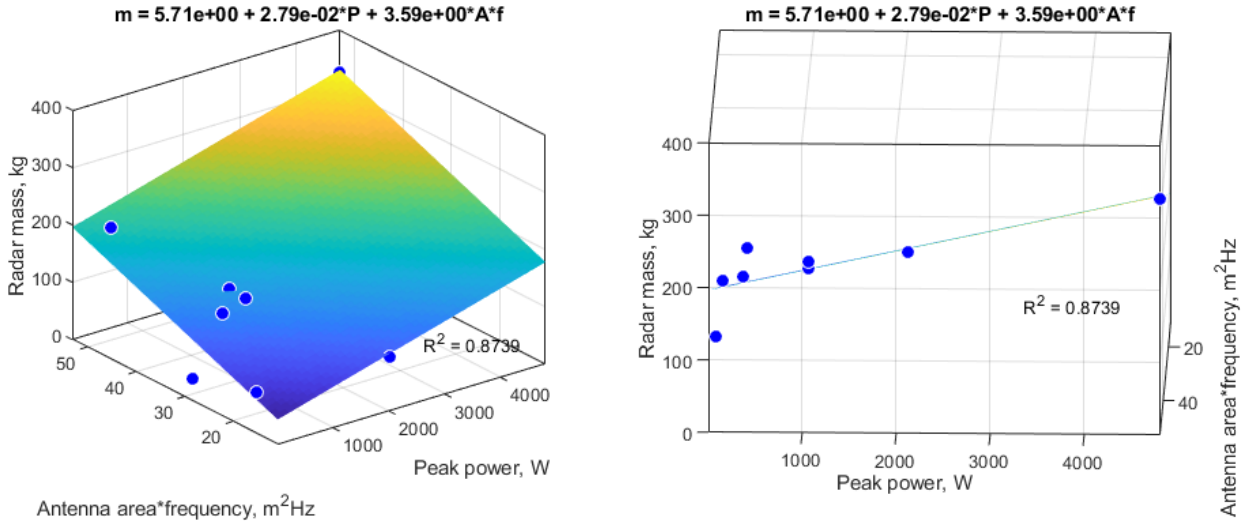
A polynomial surface was also fit to the non-scanning data set. The fit had a degree of one in the x dimension (peak power) and a degree of one in the y dimension (antenna area times frequency). The fit has an R^2 value of 0.874 and is depicted from two different points of



(a) View 1 (b) View 2

Figure 3.9: Scanning radar mass fit and data points.

view with the relevant data set in Fig. 3.10.



(a) View 1 (b) View 2

Figure 3.10: Non-scanning radar mass fit and data points.

The block diagram for the radar sizing module is shown in Fig. 3.11. The power used by the radar is assumed to have an efficiency of 60%, so the average power used by the radar is equal to $\overline{P} = P_t \text{PRF} \tau / 0.60$, where PRF is the pulse repetition frequency and τ is the

pulse width. The radar power consumption is assumed to be zero when the radar is not transmitting.

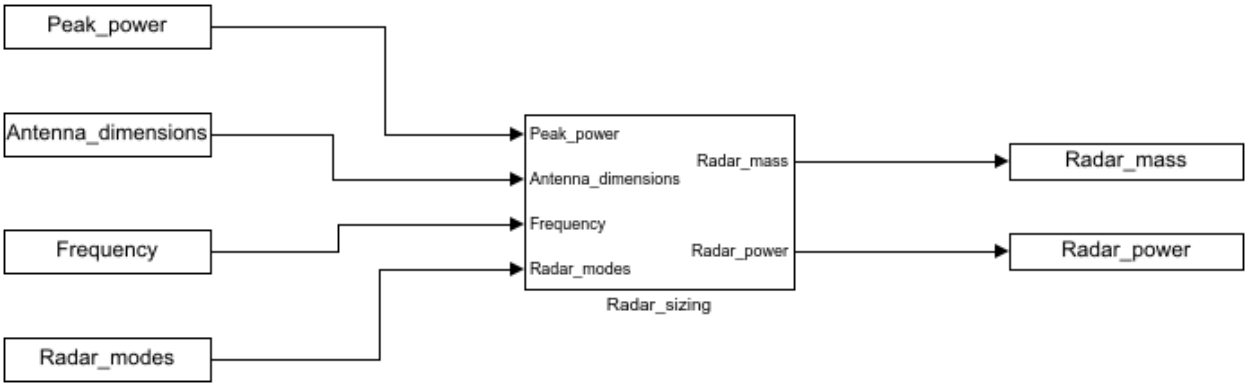


Figure 3.11: Block diagram for radar sizing module.

3.4.3 Payload: Optical

Optical payloads are sized based on analogy with existing systems, as in the original DISCO formulation. The scaling equations for optical payloads are [98]:

$$R = \frac{A_n}{A_r} \tag{3.10}$$

$$M_n \approx K R^3 M_r \tag{3.11}$$

$$P_n \approx K R^3 P_r \tag{3.12}$$

$$K = \begin{cases} 1, & R \geq 0.5 \\ 2, & R < 0.5 \end{cases} \tag{3.13}$$

where

R = area ratio
A_n = aperture of new optical payload
A_r = aperture of reference optical payload
M_n = mass of new optical payload
M_r = mass of reference optical payload

K = scaling factor
P_n = power of new optical payload
P_r = power of reference optical payload

The block diagram for the optical sizing module is shown in Fig. 3.12.

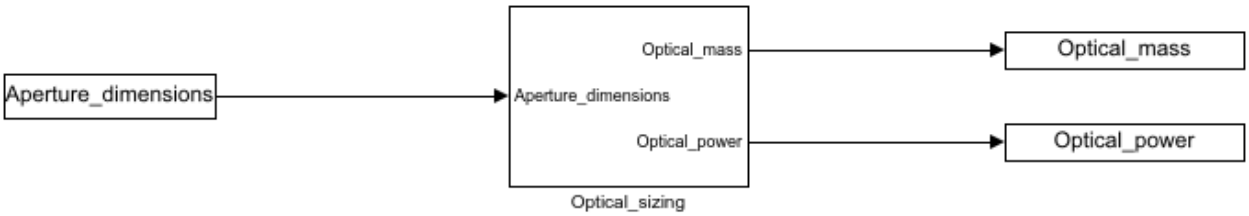


Figure 3.12: Block diagram for optical sizing module.

3.4.4 Power

The power sizing module estimates the satellite power budget and sizes the power subsystem appropriately. The block diagram is shown in Fig. 3.13.

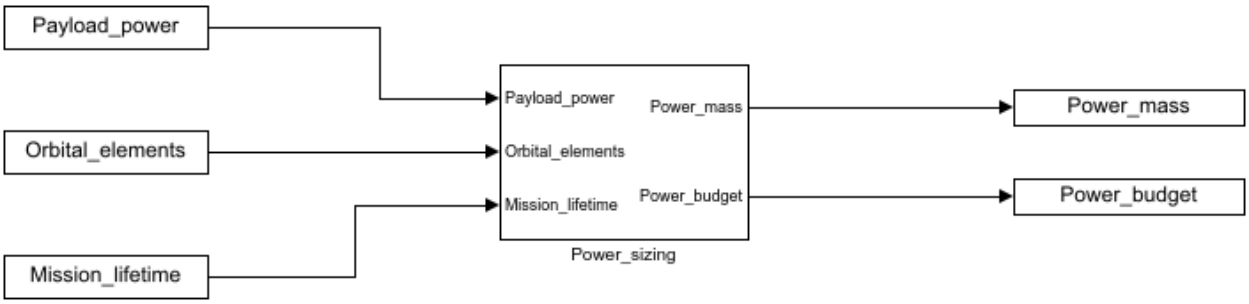


Figure 3.13: Block diagram for power sizing module.

1. To size the power subsystem, the total power requirements for the satellite must be estimated. The average power used by the payload is calculated in the payload sizing module. The average power used by the satellite as a whole can then be estimated from the payload power based on historical data as outlined in Table 14-20 of Ref. 165.

For low-Earth orbit satellites with on-board propulsion, the payload power is assumed to account for 46% of the total power.

2. The amount of time in which the satellite is illuminated and the amount of time in which the satellite is eclipsed is calculated as [165],

$$T_E = \max_{\nu \in [0, 2\pi]} \frac{2 \sin^{-1}(r_E / r(\nu))}{\dot{\nu}} \tag{3.14}$$

$$T_i = T - T_E \tag{3.15}$$

$$\begin{pmatrix} r(\nu) & = & p/(1 + e\cos(\nu)) \\ \dot{\nu} & = & na^2\sqrt{1-e^2}/r(\nu)^2 \end{pmatrix}$$

where

T_E = maximum eclipse time
T_i = minimum illumination time
T = satellite orbital period, $T = 2\pi\sqrt{a^3/\mu}$
ν = true anomaly, rad
r_E = planet radius
r = magnitude of satellite position vector
$\dot{\nu}$ = rate of change of true anomaly, rad/s
p = semiparameter, $p = a(1 - e^2)$
a = orbital semimajor axis
e = orbital eccentricity
n = orbital mean motion, $n = \sqrt{\mu/a^3}$
μ = standard gravitational parameter

3. The power needed in light versus in eclipse must be calculated. It is assumed that optical payloads only operate during illuminated periods and that radar payloads that do not operate continuously operate as much in the light as possible. As a worst-case scenario, it is assumed that communications payloads operate as much in the dark as possible. All other subsystems are assumed to require constant power. The average power needed in eclipse is P_E, and the average power needed in illumination is P_i.

4. The amount of power that must be generated during the illuminated period is calcu-

lated. Assuming the power subsystem operates at 85% efficiency during illumination ($X_i = 0.85$) and 65% efficiency during eclipse ($X_E = 0.65$), then the power generated during illumination, P_{gen} is [165]

$$P_{gen} = \frac{1}{T_i}\left(\frac{P_E T_E}{Xe} + \frac{P_i T_i}{X_i}\right) \tag{3.16}$$

5. The total energy E_E that must be stored for use during the eclipse period is calculated as

$$E_E = \frac{P_E T_E}{Xe} \tag{3.17}$$

6. The solar panel size can be estimated based on the power generation requirement. The panels are assumed to be GaAs with a production efficiency e_s of 28% and a mass per square meter ρ_A of 2.8 kg. Additionally, the system is assumed to have an inherent degradation I_d due to design, assembly, and temperature issues of 15%. The panels are also assumed to experience a degradation of 3.75% per year [165]. The total solar panel area needed, A_s, is therefore

$$A_s = P_{gen}/(e_s S I_d \cos(\theta) * 0.9625^l) \tag{3.18}$$

where l is the mission lifetime in years, S is the solar constant for the planet of interest, and θ is the angle of incidence for the light on the solar panels. θ is assumed to be 45° as a conservative estimate in the absence of sun pointing. The solar panel mass m_s is therefore $m_s = A_s \rho_A$.

7. The battery is assumed to be a lithium ion battery, with an energy density ρ_e of 125

Whr/kg and an efficiency e_b of 98%. The needed battery mass m_b is therefore

$$m_b = E_E \rho_b / e_b \tag{3.19}$$

8. The total mass of the subsystem is estimated as the sum of the panel mass, battery mass, actuator mass, and cable mass. There are assumed to be two actuators weighing 2 kg each. The cable mass is assumed to make up 25% of the subsystem, as recommended by Ref. 98. The total power mass m_p is

$$m_p = (m_s + m_b + 4\text{kg})/0.75 \tag{3.20}$$

3.4.5 Propulsion

The propulsion sizing module calculates the change in velocity (Δv) needed for station-keeping and constellation manifestation, then sizes the subsystem accordingly. Propellant properties (density, specific impulse) can be provided as an input or the default values for hydrazine will be used. Additionally, the module requires knowledge of the ballistic coefficient, $B = m/(C_d A)$. In the absence of a specified value, the drag coefficient C_d is assumed to be 2.2 [129]. The relationship between mass and area can be approximated as [11]

$$m = 106 A_{eff}^{1.04} \tag{3.21}$$

Using this equation, a 10 kg nanosatellite would expect an A_{eff}/m value of 0.0103 m^2/kg, a 500 kg small satellite a value of 0.0089 m^2/kg, and a 3,000 kg satellite a value of 0.0083 m^2/kg. The area to mass ratio is therefore conservatively estimated as 0.01 m^2/kg. These values yield a default ballistic coefficient of 45.45 kg/m^2. The propulsion sizing module

assumes a circular or near-circular orbit.

The components of the Δv budget are

- **Drag:** Air density at a given altitude a can be estimated by interpolating from density values given in the CIRA-72 model [154]. Using this model, the air density ρ is equal to

$$\rho = \rho_r e^{-(a-a_r)/H} \tag{3.22}$$

  where ρ_r is the reference density, a_r is the reference altitude, and H is the scale height provided by the model. The Δv due to drag can then be estimated as [96]

$$\Delta v_d = \rho v^2 t/(2B) \tag{3.23}$$

  where v is the spacecraft velocity and t is the mission lifetime. The analysis can be extended to elliptical orbits as

$$\Delta v_d = \frac{t}{T}\int_0^{2\pi} \rho(\nu) v(\nu)^2/(2B) d\nu \tag{3.24}$$

  where T is the orbital period and ν is the true anomaly in radians.

- **Gravitational perturbations:** In a satellite constellation, cluster, or formation of differing inclinations, semimajor axes, or eccentricities, the satellites will experience relative secular drift due to oblateness effects, primarily due to the J_2 gravitational term. This drift can be permitted to occur, introducing a long-term periodicity to coverage performance of the constellation due to the relative drift. Otherwise, the perturbations must be offset by occasional maneuvers. This analysis considers only the offset of relative drift due to the J_2 term and does not consider any higher-order zonal terms nor any of the tesseral terms because these terms are relatively small—the

next zonal term (J_4) is three orders of magnitude smaller than J_2. The relative drift in orbital elements due to higher order terms is therefore typically less than 1% of the value of the relative drift due to J_2 in the presence of inclination/semimajor axis differences, but these higher order terms can be included for completeness if desired. The orbits are assumed to be circular or near-circular.

The rates of change of RAAN ($\dot{\Omega}$) and argument of periapsis ($\dot{\omega}$) due to the J_2 term are [154]

$$\dot{\Omega} = \frac{3nJ_2r_e^2}{2a^2(1-e^2)^2}\cos(i) \tag{3.25}$$

$$\dot{\omega} = \frac{3nJ_2r_e^2}{2a^2(1-e^2)^2}(5\sin(i)^2/2 - 2) \tag{3.26}$$

The difference between these rates and that of a reference orbit is then calculated. For a cluster or formation, the reference orbit is the nominal orbit of the center of the cluster. For a constellation, the mean of the $\dot{\Omega}$ values is found and the orbit with the value closest to the mean used as the reference orbit. The relative change over an orbit is obtained by multiplying the relative rates by the orbital period. The satellite is assumed to make impulsive burns each orbit to correct the perturbations. The Δv for correcting the argument of periapsis drift is [167]

$$\Delta v_\omega = 2e\sqrt{\frac{\mu}{a(1-e^2)}}\sin(\Delta\omega/2) \tag{3.27}$$

The Δv for correcting the RAAN drift of a circular orbit is [167]

$$\Delta v_\Omega = 2v\sin(\theta/2), \tag{3.28}$$

$$\theta = \cos^{-1}\left(\cos(i)^2 + \sin(i)^2\cos(\Delta\Omega)\right) \tag{3.29}$$

The total propellant used for gravitational corrections over the course of the mission is

$$\Delta v_g = (\Delta v_\omega + \Delta v_\Omega)\frac{t}{T} \tag{3.30}$$

- **Initial spacing:** The satellites in a constellation are assumed to be placed into their final orbit by the launch vehicle. If the satellites require maneuvering to distribute around the orbit, this distribution can be accomplished either by using a small amount of propellant to temporarily raise the orbit or by differential drag. True anomaly adjustment is therefore not considered in the propellant budget, nor is initial spacing for constellations. Initial spacing for clusters and formations relative to a circular orbit is included in the budget. Initial RAAN spacing is achieved according to Eq. 3.29. A Hohmann transfer is used to adjust the semimajor axis and eccentricity. It is assumed that the satellites have no inclination separation. The Δv needed for initial spacing is termed Δv_i.

- **Additional maneuvers:** If the Δv for any additional maneuvers is known, it is added into the propellant budget as Δv_m.

- **Margin:** Some margin is added to the Δv budget. Unless specified, the margin M_f is assumed to be 25%.

The total Δv is equal to $\Delta v_T = (\Delta v_d + \Delta v_g + \Delta v_i + \Delta v_m)/(1 - M_f)$. The satellite dry mass without the propulsion subsystem, m_{d0} is estimated based on the payload mass and power mass as described in the following section. The engine mass m_e is assumed to be 8.51 kg, the mass of the Aerojet MR80B engine. The tank mass is approximated as [165]

$$m_T(m_p) = 27.086(m_p/\rho)^3 - 61.703(m_p/\rho).^2 + 66.29(m_p/\rho) + 1.3192 \tag{3.31}$$

where the tank mass m_T and propellant mass m_p are in meters and the propellant density is in kg/m^3. The rocket equation is rearranged to form

$$m_p = (m_{d0} + m_e + m_T(m_p))(e^{-\Delta v_T/(I_{sp}g0)} - 1) \tag{3.32}$$

where I_{sp} is the specific impulse and $g_0 = 9.81\text{m/s}^2$. Eq. (3.4.5) is then solved iteratively to obtain the propellant mass. The total propulsion subsystem mass (not including propellant) is the sum of the engine mass and the tank mass.

The block diagram for the propulsion sizing module is shown in Fig. 3.14.

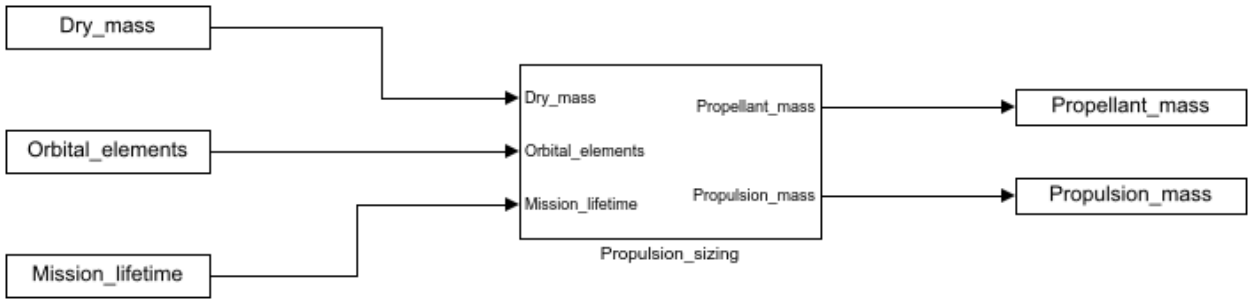


Figure 3.14: Block diagram for propulsion sizing module.

3.4.6 Other

For subsystems not explicitly described above, DISCO-Tech does not use a dedicated sizing module. When using existing subsystems with known masses, these masses can be included in the subsystem definition. For subsystems that are not pre-existing, historical mass fractions are used [165]. The payload mass is used to estimate the total dry mass, then the mass fractions are used to estimate the attitude determination and control system, structure, thermal, processing, TT&C (for satellites with no communications payload), and miscellaneous masses. Crosslinking systems are not modelled, as cluster missions are assumed to have

small inter-satellite distances capable of using low-mass, low-power RF systems or small laser crosslinks. The block diagram for the miscellaneous sizing module is shown in Fig. 3.15.

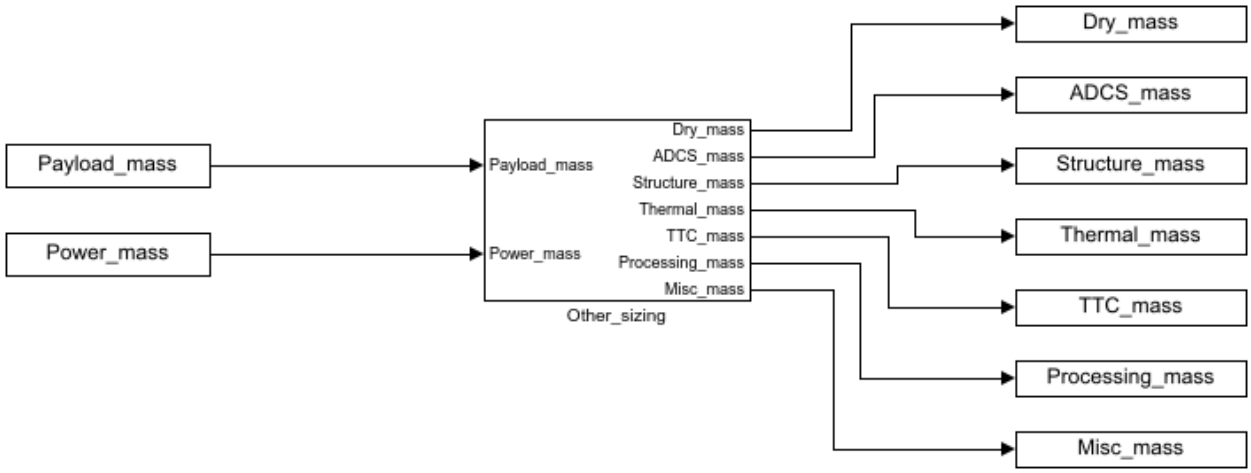


Figure 3.15: Block diagram for miscellaneous sizing module.

An example block diagram for the entire satellite sizing process is shown in Fig. 3.16, assuming radar and communications payloads.

3.5 Constraints

As mentioned in subsection 3.2.3, constraints can be handled either through repair or through penalty functions. Typical DISCO-Tech constraints are described below. DISCO-Tech also accepts custom constraints.

3.5.1 Link budget

The sizing process for the spacecraft described in subsection 3.4.1 typically ensures that the link budget is satisfied by adjusting the transmit power until it is sufficient to provide the

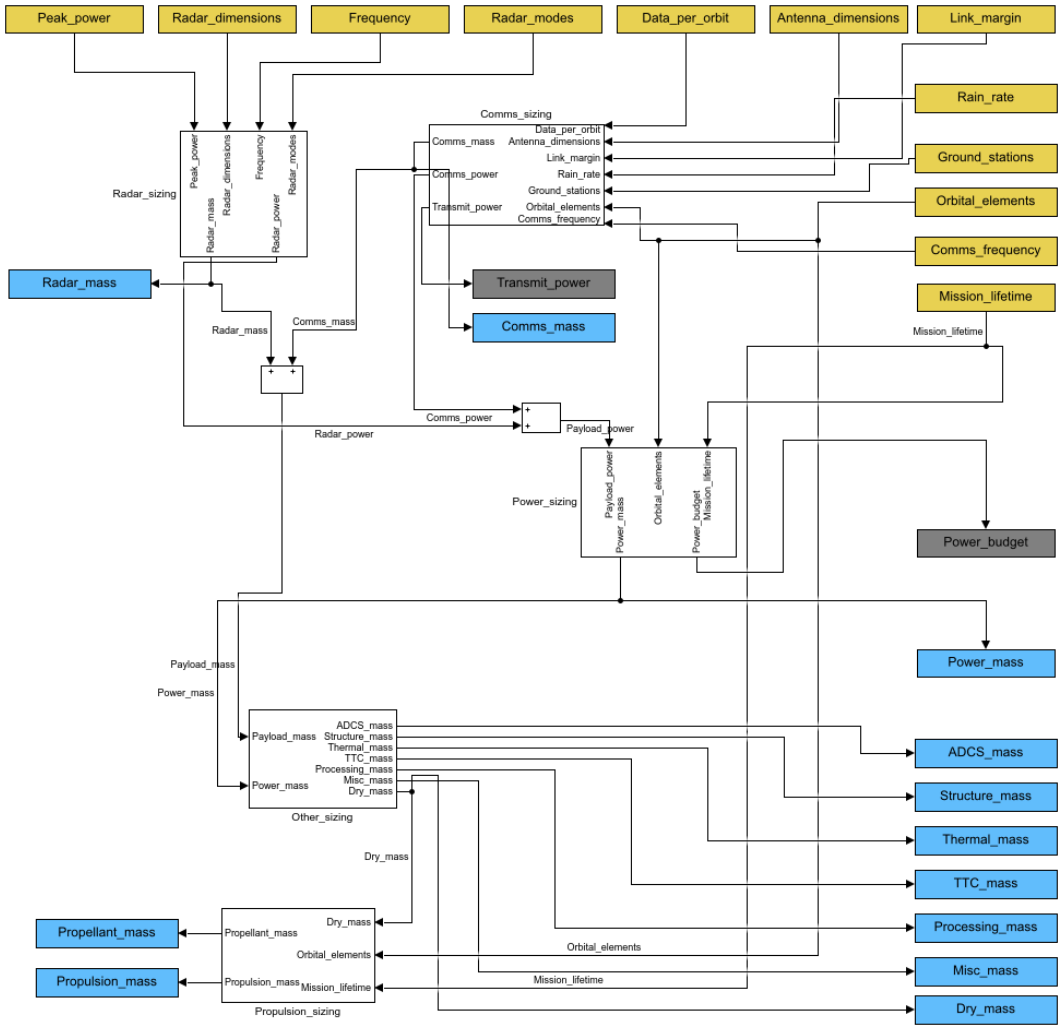


Figure 3.16: Block diagram for sample satellite sizing module.

desired bit error rate or E_b/N_0. Alternately, a link budget constraint can be imposed when the transmit power is limited or when it is desired to adjust other parameters to satisfy the link budget. For a COTS communications subsystem, the inclusion of the constraint is necessary since the transmit power cannot be adjusted. The constraint is of the form

$$g(\mathbf{x}, \mathbf{y}) = BER(\mathbf{x}, \mathbf{y}) - BER_{des} \leq 0 \tag{3.33}$$

where $\mathbf{x}$ and $\mathbf{y}$ are decision variable vectors containing information about the transmit power, frequency, path length, data rate, and other link budget parameters. $BER(\mathbf{x},\mathbf{y})$ is the achievable bit error rate for the system. BER_{des} is the desired bit error rate.

If the link budget constraint is violated, the violation can be repaired by altering one or more parameters that affect the link budget, including altitude, transmit power, frequency, antenna diameter, and data rate.

An example penalty function for a link budget constraint would be:

$$p(\mathbf{x},\ \mathbf{y}) = a\left(\log_{10}\left(BER(\mathbf{x},\ \mathbf{y})\right) - \log_{10}\left(BER_{des}\right)\right) + b \tag{3.34}$$

where a is a scaling coefficient and b is an offset coefficient.

3.5.2 Power budget

The sizing process also ensures that the power budget is satisfied. Alternately, a power budget constraint can be imposed when power generation or storage capabilities are limited, such as when using COTS power components or fixed solar panel dimensions, or when power subsystem materials are decision variables. If the power budget is violated, the violation can be repaired by altering one or more parameters that affect the power budget, including battery size, battery material, solar panel size, solar panel material, payload duty cycle, and any other parameters affecting payload power usage.

An example penalty function for a power budget constraint would be:

$$p(\mathbf{x},\ \mathbf{y}) = a(P_n(\mathbf{x},\ \mathbf{y}) - P_{gen}(\mathbf{x},\ \mathbf{y})) + b \tag{3.35}$$

where P_n is the power needed by the satellite, P_{gen} is the power generated by the satellite,

a is a scaling coefficient, and b is an offset coefficient.

3.5.3 Mass/count

An upper limit can be imposed on the maximum mass of the satellite. Due to the fact that the mass in influenced by many decision variables, a penalty function is employed to limit the mass. If the mass is excessive, the solution is infeasible and returns the penalty function value in lieu of the objective value. The mass constraint can be imposed on either a per-satellite or per-orbit basis, the latter case ensuring that the mass limit for some launch vehicle is satisfied when that vehicle is used to launch all satellites in a particular orbit.

An example penalty function for a mass constraint would be:

$$p(\mathbf{x},\ \mathbf{y}) = b + a \sum_{i=1}^{n_{sats}} \min(m_i(\mathbf{x},\ \mathbf{y}) - m_{lim}, 0) \tag{3.36}$$

where m_i is the mass of the i-th satellite, m_{lim} is maximum satellite mass, a is a scaling coefficient, and b is an offset coefficient.

Bounds can also be imposed on the total number of satellites in the system. This constraint is typically repairable: If the satellite count limit is violated, the violation can be repaired by adjusting the number of satellites per plane or number of planes until the limit is satisfied. Alternately, bounds can also be imposed on the total instances of a particular payload in the system. This constraint is typically repairable. If the payload count limit is violated, the violation can be repaired by adjusting the number of satellites per plane, the number of planes, or the satellites to which the payload is assigned until the limit is satisfied. Such a constraint could be written as

$$l \leq \sum_{i=1}^{n_{sats}} \delta_i(\mathbf{x}) \leq u \tag{3.37}$$

If the limit is on the number of payloads, δ_i is one if i-th satellite has the payload and zero otherwise. If the limit is on the number of satellites, δ_i is one. l is the lower limit on the number of payloads/satellites and u is the upper limit on the number of payloads/satellites. An example penalty function for a constraint on the number of payloads would be:

$$p(\mathbf{x}) = b + a \max\left(l - \sum_{i=1}^{n_{sats}} \delta_i(\mathbf{x}), -u + \sum_{i=1}^{n_{sats}} \delta_i(\mathbf{x}), 0\right) \tag{3.38}$$

where a is a scaling coefficient and b is an offset coefficient.

3.5.4 Repeating ground track orbits

Repeating ground track orbits are widely used, particularly in scientific missions where it is desirable to repeatedly cover the same ground locations to compare measurements over time. A ground track is repeating if the satellite undergoes an integer number of orbits during an integer number of "days", where a day here refers to the amount of time for a spot on the ground to return to the same position relative to the satellite's orbit. This definition accounts for the effects of orbital precession and is called the nodal period of the ground point. If ω_P is the angular speed of the planet, the nodal period of a ground point is $T_G(a, e, i) = 2\pi/(\omega_P - \dot{\Omega}(a, e, i))$. $\dot{\Omega}(a, e, i)$ is the rate of change of the satellite's RAAN due to oblateness effects, as defined in Section 3.3. The nodal period of the ground point is a function of the satellite's semimajor axis, eccentricity, and inclination. Since the planet's oblateness also affects the position of the satellite's ascending node, the orbital period must also be adjusted to compensate. The nodal period of the satellite is therefore $T_S(a, e, i) = 2\pi/(n(a) + \dot{M}_0(a, e, i) + \dot{\omega}(a, e, i))$, where $n(a)$ is the satellite mean motion and $\dot{M}_0(a, e, i)$ and $\dot{\omega}(a, e, i)$ are the rates of change of the reference mean anomaly and the argument of periapsis due to oblateness effects, as described in Section 3.3.

The orbit is a repeating ground track orbit if there exist integers m and$n$ such that

$$mT_S(a, e, i) = kT_G(a, e, i) \tag{3.39}$$

k can be fixed to require that the satellite repeat in a particular number of days. The current values for a, e, and i, are provided by the optimizer or as inputs. If the value for m that solves Eq. (3.39) is not integer, then the constraint is violated. Due to the difficulty of finding periods that cause the integrality requirement for m to be exactly satisfied, this constraint is always implemented as a repairable constraint. If the constraint is violated, the calculated value for m, m_c, is rounded to the nearest integer. By solving

$$\lfloor m_c \rceil T_S(a, e, i) = kT_G(a, e, i) \tag{3.40}$$

for a, the nearest orbit that satisfies the repeating ground track constraint is found. $\lfloor \cdot \rceil$ denotes the rounding operation. The genome is repaired by replacing the satellite's semimajor axis with the calculated value.

3.5.5 Fuel/Reachability

Satellites are limited in the amount of propellant they can carry. If a maneuver is scheduled through the optimization, the maneuver most be possible given the amount of propellant available on the satellite. This constraint can be written as

$$g(\mathbf{x}, \mathbf{y}) = \Delta v(\mathbf{x}, \mathbf{y}) - \Delta v_{lim} \leq 0 \tag{3.41}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the decision variable vectors containing information about a satellite's orbit, maneuver, mass, and propulsion properties. $\Delta v(\mathbf{x}, \mathbf{y})$ is the change in velocity needed

to complete the maneuver. Δv_{lim} is the change in velocity that is achievable given the amount of propellant available for the maneuver. The amount of Δv required for a maneuver must be calculated in order to see if Eq. (3.41) is satisfied. If Eq. (3.41) is satisfied for a combination of initial and final orbits when considering limitations such as the maximum thrust, the final orbit is said to be reachable from the initial orbit. This constraint is typically repairable.

Previous papers on constellation reconfiguration have restricted reachability analysis to specific sets of maneuvers. One study restricted reconfiguration to in-plane maneuvers, then used a genetic algorithm to solve for the two-burn transfers yielding the best coverage in the final configuration [53]. Other studies restrict the initial and final constellations to known sets of orbits, presolving for the propellant needed to go between each combination of orbits and then solving the assignment problem to find the optimal set of transfers [10, 51, 100].

If a maneuver is specified by the Cartesian components of the instantaneous velocity change vector $\Delta\mathbf{v}$, the total Δv needed for the maneuver is $\|\Delta\mathbf{v}\|$. If there is insufficient propellant to perform the desired maneuver, the magnitude of the maneuver is reduced so it is achievable. However, this method limits the number of burns, though the variable-length GA formulation can offset this limitation. Furthermore, convergence is difficult due to the dependence of the maneuver's impact on the orbital elements, time at which the maneuver occurs, and previous maneuvers. This method also assumes instantaneous transfers.

A general framework for rapidly determining the reachability of one orbit from another for low-thrust systems is desired. Although methods exist for generating the reachable set [77, 78, 171, 174], they rely on numerical simulation and are too computationally expensive to call for each solution during the optimization, since the reachable set will change as the initial orbit changes. Instead, an estimation of the propellant used is generated using a linearized version of Gauss's Variational Equations (GVE). If the estimated propellant is less than or equal to the available propellant, the final orbit is considered to be reachable

from the initial orbit. This section outlines the linearization and examines the regions for which the linearization is within some user-defined bounds of the nonlinear system.

Previous research linearized GVE about the final orbit and used the resulting equations with model predictive control (MPC) to calculate the required controls to maneuver from one orbit to another [19]. This linearization serves as the basis for the approach described here, though the MPC process was deemed unnecessarily costly. Only the total propellant expenditure is needed, not the entire control history. Furthermore, the linearization presented in Ref. 19 is improved upon through the use of the modified equinoctial orbital elements, by treating the true longitude as an independent parameter, and by improved analysis of the validity of the linearization. It is assumed that the final value of the true longitude is irrelevant, as it can be set afterward by temporarily raising or lowering the orbit using a comparatively small amount of propellant or by holding the satellite at an intermediate stage in its orbit until the desired phasing has been reached. It is also assumed that the maximum acceleration of the spacecraft does not change over time despite the change in the spacecraft's mass.

The version of the GVE employed uses the set of modified equinoctial orbital elements (MEOE) and control accelerations in the local vertical, local horizontal (LVLH) frame. These equations are [161]:

$$\begin{bmatrix} \dot{p} \\ \dot{f} \\ \dot{g} \\ \dot{h} \\ \dot{k} \\ \dot{L} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \sqrt{\mu p}\frac{q^2}{p^2} \end{bmatrix} + \begin{bmatrix} 0 & \sqrt{\frac{p}{\mu}}\frac{2p}{q} & 0 \\ \sqrt{\frac{p}{\mu}}\sin(L) & \sqrt{\frac{p}{\mu}}\frac{(q+1)\cos(L)+f}{q} & -\sqrt{\frac{p}{\mu}}\frac{g(h\sin(L)-k\cos(L))}{q} \\ -\sqrt{\frac{p}{\mu}}\cos(L) & \sqrt{\frac{p}{\mu}}\frac{(q+1)\cos(L)+g}{q} & \sqrt{\frac{p}{\mu}}\frac{f(h\sin(L)-k\cos(L))}{q} \\ 0 & 0 & \sqrt{\frac{p}{\mu}}\frac{(1+h^2+k^2)\cos(L)}{2q} \\ 0 & 0 & \sqrt{\frac{p}{\mu}}\frac{(1+h^2+k^2)\sin(L)}{2q} \\ 0 & 0 & \sqrt{\frac{p}{\mu}}\frac{h\sin(L)-k\cos(L)}{q} \end{bmatrix} \begin{bmatrix} u_r \\ u_\theta \\ u_h \end{bmatrix} \tag{3.42}$$

where

u_r = control acceleration in the radial direction (LVLH)
u_θ = control acceleration in the tangential direction (LVLH)
u_h = control acceleration in the angular momentum direction (LVLH)
L = true longitude
e = eccentricity
p = semiparameter
i = inclination
Ω = RAAN
ω = argument of periapsis
$f = e\cos(\omega + \Omega)$
$g = e\sin(\omega + \Omega)$
$h = \tan(\frac{i}{2})\cos(\Omega)$
$k = \tan(\frac{i}{2})\sin(\Omega)$
$q = 1 + f\cos(L) + g\sin(L)$

More concisely, the equations can be written as

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f(x)} + B(\mathbf{x})\mathbf{u} \tag{3.43}$$

where

$\mathbf{x}$ = $[p, f, g, h, k, L]^T$
$\mathbf{u}$ = $[u_r, u_\theta, u_h]^T$
$\mathbf{f(x)}$ = $\in \mathbb{R}^6$, represents the dynamics of the elements in the absence of control
$B(\mathbf{x})$ = $\in \mathbb{R}^{6\times 3}$, input effect matrix.

L is the only fast-varying orbital element in the set. Therefore, the other orbital elements are constant in the absence of perturbations like oblateness effects. A reduced set of elements is defined as $\mathbf{z} = [p, f, g, h, k]^T$. Eq. (3.43) can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{z}\\ L\end{bmatrix} = \mathbf{f}(\mathbf{z}, L) + B(\mathbf{z}, L)\mathbf{u} \tag{3.44}$$

Since the first five elements of $\mathbf{f(x)}$ are zero for the two body problem, the growth of these

elements can be written as

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = \overline{B}(\mathbf{z}, L)\mathbf{u} \tag{3.45}$$

where $\overline{B}(\mathbf{z}, L)$ is the first five rows of $B(\mathbf{z}, L)$. The equation is affine in $\mathbf{u}$ but nonlinear in $\mathbf{z}$ due to the dependence of $\overline{B}$ on $\mathbf{z}$.

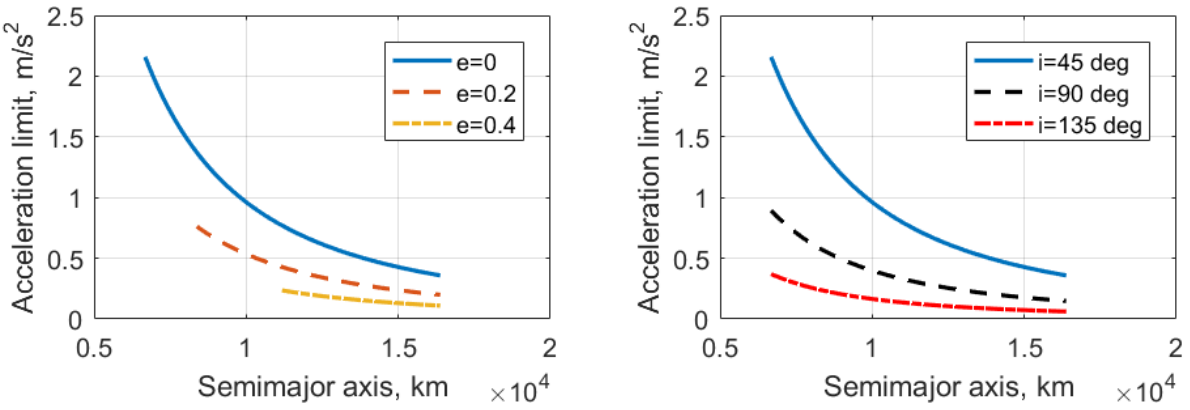
It is advantageous to treat the true longitude L as an independent parameter, so the conditions under which such an assumption holds are examined. The dynamics of L are described by

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \sqrt{\frac{\mu}{p^3}}(1 + e\cos(\nu))^2 + \sqrt{\frac{p}{\mu}}\frac{\tan(i/2)\sin(\nu + \omega)}{1 + e\cos(\nu)}u_h \tag{3.46}$$

The classical orbital elements are used to provide a sense of physical understanding, with θ being the true anomaly. The control has the greatest influence on the dynamics of L when the second term of the previous equation is maximized. Define R as the ratio between the second term in Eq. (3.46) and the first term in Eq. (3.46). R is maximized if $\sin(\theta + \omega) = 1$, $\cos(\theta) = -1$, and u_h is equal to its maximum possible acceleration $u_{\max}$,

$$R_{\max} = \frac{u_{\max}a^2(1 - e^2)^2\tan(i/2)}{\mu(1 - e)^3} \tag{3.47}$$

This ratio quantifies the maximum relative impact that the control thrust can have on the dynamics of L. By setting an upper limit on the instantaneous value of the ratio, a maximum allowable thrust acceleration for various combinations of a, i, and e can be calculated. Note that Eq. (3.47) is zero for an inclination of zero, since thrusting in the angular momentum direction at this inclination will not affect the true longitude. Figs. 3.17a and 3.17b show the maximum allowable thrust acceleration such that the ratio $R_{\max}$ at apoapsis will not exceed 0.1. When the control accelerations are smaller than those described in the plots for a given combination of semimajor axis, eccentricity, and inclination, the impact of the control on the growth of the true longitude is ignored.



(a) Thrust acceleration limit for 45° inclined orbits at various eccentricities.

(b) Thrust acceleration limit for circular orbits at various inclinations.

Figure 3.17: Thrust limits.

Nanosatellites tend to employ micropropulsion systems with low thrusts on the order of millinewtons [104]. The control influence on L can be safely ignored for these cases. High-thrust CubeSat propulsion systems with up to 1.25 N of thrust are in development[2]. Assuming that these systems would be used exclusively with larger CubeSats with a dry mass of at least 5 kg, the corresponding acceleration would be at most 0.25 m/s^2. From Figs. 3.17a and 3.17b, it can be seen that satellites in LEO with eccentricity less than 0.4 and inclinations less than 135° will satisfy this constraint. If a nanosatellite is constrained to LEO, either due to rideshare availability or due to sensor requirements, the assumption that the control does not significantly affect the true longitude will hold. The rate of change of L is then $\mathrm{d}L/\mathrm{d}t = \sqrt{\mu p}q^2/(p^2)$. The values of L can be approximated either by holding the MEOE fixed at either the initial or final values or by linearly interpolating between the initial and final values and calculating the growth of L at each time step. The approximation of the secular growth rate distorts the relationship between time steps and steps of L, but the validation cases tested in subsubsection 3.5.5 demonstrate that the effect does not

[2] http://www.rocket.com/files/aerojet/documents/CubeSat/MPS-130%20data%20sheet%20crop.pdf

significantly impact the results. The effect can be mitigated by using smaller time steps or by iterating through the problem multiple times, using the previous solution to generate the approximation of L. The initial formulation merely linearly interpolates values of a and e between the initial and final values over the analysis period, then uses those values to calculate the true longitude. Using this process, true longitude can be treated as a function of time alone, allowing it to be treated as an independent parameter in the linear program if the control thrusts are within the bounds specified in Figs 3.17a and 3.17b. Eq. (3.45) can then be rewritten as $d\mathbf{z}/dt = \overline{B}(\mathbf{z}, L(t))\mathbf{u}$.

Performing a Taylor series expansion of Eq. (3.45) about some reference orbit $\mathbf{z}_s$ gives

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = \left(\overline{B}(\mathbf{z}_s, L(t)) + \left.\frac{\partial \overline{B}}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{z}_s} \Delta\mathbf{z}\right)\mathbf{u} + \mathrm{HOT} \tag{3.48}$$

where $\Delta\mathbf{z} = \mathbf{z} - \mathbf{z}_s$. $\partial\overline{B}/\partial\mathbf{z}$ in the second term is a tensor of rank three. The Higher Order Terms (HOT) are neglected. The equation for a single element z_i is given below for clarity, with b_{ij} being the element of B in the i-th row and the j-th column.

$$\frac{\mathrm{d}z_i}{\mathrm{d}t} \approx \begin{bmatrix} b_{i1} & b_{i2} & b_{i3} \end{bmatrix}\mathbf{u} + \begin{bmatrix} \Delta p \\ \Delta f \\ \Delta g \\ \Delta h \\ \Delta k \end{bmatrix}^T \left.\begin{bmatrix} \partial b_{i1}/\partial p & \partial b_{i2}/\partial p & \partial b_{i3}/\partial p \\ \partial b_{i1}/\partial f & \partial b_{i2}/\partial f & \partial b_{i3}/\partial f \\ \partial b_{i1}/\partial g & \partial b_{i2}/\partial g & \partial b_{i3}/\partial g \\ \partial b_{i1}/\partial h & \partial b_{i2}/\partial h & \partial b_{i3}/\partial h \\ \partial b_{i1}/\partial k & \partial b_{i2}/\partial k & \partial b_{i3}/\partial k \end{bmatrix}\right|_{\mathbf{z}=\mathbf{z}_s} \mathbf{u} \tag{3.49}$$

$$\approx (\mathbf{b}_i + (J_i\Delta\mathbf{z})^T)\mathbf{u} \tag{3.50}$$

The second term results in a nonlinear equation, since the term contains a product of $\Delta\mathbf{z}$ and $\mathbf{u}$. The system can be solved using linear programming techniques if the nonlinear term can be neglected. This simplification is valid if the components of the nonlinear term $(J_i\Delta\mathbf{z})$ are

much smaller than the components of the linear term ($\mathbf{b}_i$). The ratio of the magnitudes of the nonlinear and linear terms causing a change in element i due to the difference in element j from the stationary orbit is

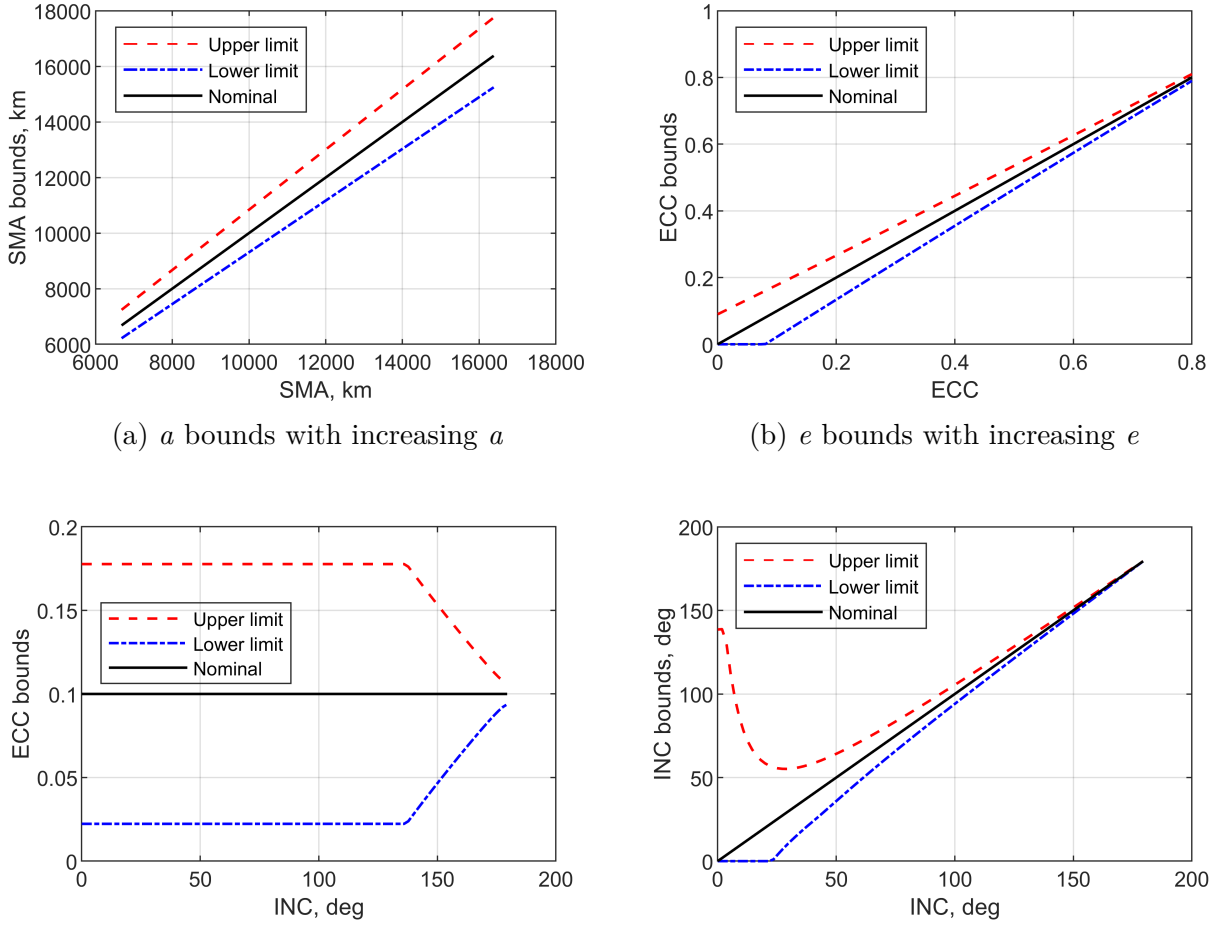
$$Q_{ij} = \frac{|\Delta z_j| \|\mathbf{j}_{ij}\|}{\|\mathbf{b}_i\|} = \frac{|\Delta z_j| \sqrt{(\partial b_{i1}/\partial z_j)^2 + (\partial b_{i2}/\partial z_j)^2 + (\partial b_{i3}/\partial z_j)^2}}{\sqrt{b_{i1}^2 + b_{i2}^2 + b_{i3}^2}} \tag{3.51}$$

$\mathbf{j}_{ij}$ is the j-th column of the J_i matrix. By setting an upper limit on the value of Q_{ij}, $Q_{\max}$, bounds on each of the orbital elements can be developed. Since the bounds on Δz_j must satisfy the imposed limits on Q_{ij} $\forall i \in [1, 5]$, the smallest value of Δz_j calculated by the five equations is selected by taking the minimum over i. This limit should be satisfied for all values of L. However, some elements of $\overline{B}$, such as b_{21}, go to zero at certain values of L, and the ratio near these points is poorly defined. Furthermore, a large Q_{ij} value occurring when $\mathbf{b}_i$ is small still results in a small magnitude change in $\mathbf{b}_i$. To avoid these singularities, the denominator used in Eq. (3.51) is not the instantaneous value for a given L but the average value calculated by averaging $\|\mathbf{b}_i\|$ over L. This average value is denoted as $\|\mathbf{b}_i\|_{\text{avg}} = 1/(2\pi) \int \|\mathbf{b}_i(\mathbf{z}_s, L)\| dL$. Since all terms in the numerator of Eq. (3.51) are a function of L, the minimum of $|\Delta z_j|$ is taken over L to ensure the most conservative bounds. Mathematically, this relationship can be written by inverting Eq. (3.51) to obtain

$$|\Delta z_j| \leq \min_{i \in [1,5]} \left(\min_{L \in [0, 2\pi]} \left(Q_{\max} \frac{\|\mathbf{b}_i\|_{\text{avg}}}{\|\mathbf{j}_{ij}(L)\|} \right) \right) \tag{3.52}$$

The process above is used with a value of $Q_{\max} = 0.1$ to generate bounds on the linearization for an orbit with nominal values of $a = 7,000$km, $e = 0.1$, $i = \pi/4$, $\Omega = \pi/6$, and $\omega = \pi/12$. To determine the impact of the initial value of each orbital element on the bounds, the initial orbital elements are varied one at a time while holding the others fixed. a is varied from 6,678 km to 16,378 km, e is varied from 0 to 0.8, i is varied from 0 to 180°, Ω is varied from

0 to 360°, and ω is varied from 0 to 360°. Fig. 3.18 shows the results for the combinations of orbital parameters having the most significant impacts.



(a) a bounds with increasing a

(b) e bounds with increasing e

(c) e bounds with increasing i

(d) i bounds with increasing i

Figure 3.18: Element limits for valid linearization under changing initial element values.

The main determining factor in the semimajor axis bounds is itself, with larger semimajor axes having larger bounds. For a low Earth orbit, a limit of $|\Delta a| = 500$ km ensures the bounds on the linearization are satisfied, as is shown in Fig. 3.18a. These bounds only consider the GVE linearization and do not account for drag or other nonlinearities. Similarly, eccentricity is the main determining factor for its own bounds with larger eccentricities having smaller bounds, though inclination influences the bounds as well, as shown in Figs. 3.18b and 3.18c.

For an orbit with low eccentricity, bounds of about 0.1 are acceptable. As shown in Fig. 3.18d, the inclination bounds decrease with increasing inclination. A high-thrust CubeSat propulsion system may provide approximately $\Delta v = 570$ m/s for a 5 kg satellite. The achievable inclination change with $\Delta v = 570$ m/s at an altitude of 1,000 km is only 4.4°. This high propellant requirement, coupled with the large bounds on inclination, make it unlikely that a satellite would maneuver more than a couple of degrees in inclination, rendering the linearization valid for all practical cases barring a retrograde orbit with inclination greater than 120°.

Overall, the linearization will hold when the semimajor axis error is kept below 500 km, the eccentricity error kept below 0.1 for low eccentricities and below 0.05 for eccentricities near 0.5, the eccentricity kept below 0.5, the inclination kept below 120°, changes in AoP kept below 50°, and changes in RAAN and inclination kept below 10° for prograde orbits and 5° for retrograde orbits.

Now that it has been shown that the nonlinear term in Eq. (3.48) can be neglected in the cases described above, the linearized equation

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = \overline{B}(\mathbf{z}_s, L(t))\mathbf{u} \tag{3.53}$$

can be used to approximate the change in the orbital elements over time for cases satisfying the bounds on thrust and orbital element error described previously. As an initial guess, the reference orbit is a linear interpolation between the initial and final orbits. Fidelity can be improved by performing multiple iterations, using the solution from the previous iteration as the reference orbit for the next.

Applying the variation of constants formula to the above equation gives

$$\mathbf{z}(t) = \mathbf{z}_0 + \int_{t_0}^{t} \overline{B}(\mathbf{z}_s, L(v))\mathbf{u}(v)dv \tag{3.54}$$

If the system is discretized by setting $t = k\Delta t$ (where k is the integer describing the time step), using it to calculate $L_k = L(k\Delta t)$, and treating $\overline{B}$ and $\mathbf{u}$ as being fixed at each time step, the equation becomes

$$\mathbf{z}_{k+1} = \mathbf{z}_0 + \Delta t \sum_{j=0}^{k} \overline{B}_j \mathbf{u}_j \tag{3.55}$$

With this equation, the problem of finding a propellant-optimal trajectory between the initial and final orbits can now be formulated as a linear program (LP). If the three control thrusts are provided independently using three separate thrusters, the total Δv required for the maneuver is $\Delta v = \sum_k \Big(|u_{rk}| + |u_\theta k| + |u_{hk}|\Big)\Delta t$. However, in the case in which a single thruster is providing control, the required Δv is $\Delta v = \sum_k \sqrt{u_{rk}^2 + u_{\theta k}^2 + u_{hk}^2}\Delta t$. Even in the latter case, the simplifying assumption is made to use the former value for Δv. To treat the absolute values in the Δv formulation, it is necessary to create separate variables u_{ik}^+ and u_{ik}^-, both in $[0,u_{\max}]$, such that $u_{ik} = u_{ik}^+ - u_{ik}^-$. The minimization will ensure that at most one of u_{ik}^+ and u_{ik}^- is nonzero. The LP formulation can be written as

Minimize $\quad \Delta t \sum_k \sum_i u_{ik}^+ + u_{ik}^-$

with decision variables :

$u_{ik}^+ \quad \in [0, u_{\max}] \quad \forall\ k = 0, 1, \ldots, k_f,\ i\ \in\ \{r, \theta, h\}$

$u_{ik}^- \quad \in [0, u_{\max}] \quad \forall\ k = 0, 1, \ldots, k_f,\ i\ \in\ \{r, \theta, h\}$

such that :

$\mathbf{z}_{fd} - \zeta_f \quad \leq \mathbf{z}_0 + \Delta f \sum_{j=0}^{k_f} \overline{B}_j \mathbf{u}_j \leq \mathbf{z}_{fd} + \zeta_f$

where ζ_f is the allowable error in the final state and the desired final state is $\mathbf{z}_{fd}$. Similar LP formulations have been used for trajectory optimization using the Hill-Clohessy-Wiltshire equations or the Gim-Alfriend equations as a basis, such as in Ref. 149. Due to the fact that the control is held constant during each step, reaching an exact state may be impossible, requiring the inclusion of error bounds. Open source or commercial LP solvers can be used to obtain the optimal trajectory. Once the optimal solution is known, the orbital elements at each step k can be calculated using Eq. (3.55). The approximate Δv required is equal to the objective function value. If that value exceeds the amount achievable with the on-board propellant, the change in orbital element changes is reduced. The process is repeated until the current set of orbital element changes yields a reachable orbit.

This formulation does not consider uncertainties in position and velocity or pointing accuracy. For Earth-based systems with low-cost GPS receivers, position and velocity errors are 10 m and 25 cm/s respectively[3]. These errors are small compared to the magnitude of the position and velocity vectors and therefore have negligible impact on the orbital elements. A satellite with a pointing accuracy of 1° would experience less than a 2% change in the thrust vector from its nominal value[4]. Techniques for handling state errors in satellite maneuvering include Kalman filtering [48], model predictive control [149], and the Lyapunov direct stability theorem [136]. Similar measures could be adapted at the cost of increased computational complexity.

Validation

Two scenarios are used to test the problem formulation. First, the problem of raising a circular orbit from 1,000 km in altitude to 1,500 km in altitude using a maximum acceleration

[3] https://www.cubesatshop.com/product/nss-gps-receiver/

[4] https://www.bluecanyontech.com/static/datasheet/BCT_DataSheet_Components_StarTrackers.pdf

of 0.01 m/s^2 is considered. According to Ref. 168, the optimal low thrust orbit raise for a circular orbit is a continuous thrust in the velocity vector direction and consumes a total Δv of $\Delta v = \left|\sqrt{\mu/a_0} - \sqrt{\mu/a_f}\right|$. The time needed to complete the maneuver is $t_f - t_0 = (\mu/u_{\max})\left|a_0^{-1/2} - a_f^{-1/2}\right|$. For the given problem, the required Δv according to the equation is 237.1 m/s.

The scenario was permitted to run for 1.05 times the predicted time needed to complete the maneuver. The semimajor axis was assumed to vary linearly between the initial and final values, with these interpolated values being used in the B_k matrices at each time step. The estimated velocity required for the orbit raise is 235.9 m/s, resulting in a deviation from the analytical solution of less than 1%.

The second scenario requires an inclination change of two degrees within forty orbits. The orbit is circular and has an altitude of 1,000 km. From GVE, changes in inclination are caused by thrusting in the angular momentum direction, $\mathrm{d}i/\mathrm{d}t = ru_h \cos(\omega + \theta)/h$ [154]. Due to the cosine term, the impact from u_h is maximized at the ascending and descending nodes. Near the nodes, thrust in the angular momentum has little impact on other orbital elements, so the impact on i can be considered independently. If thrusting is performed for some time period ΔT centered on the nodal crossing with a constant acceleration u_c, the change in inclination for a single maneuver for a circular orbit can be shown to be

$$\delta i = \frac{2u_c a^2 \sin(n\Delta T/2)}{\mu\sqrt{1-e^2}} \tag{3.56}$$

where n is the mean motion. Since forty orbits are allowed for the inclination to occur, there are eighty nodal crossings and therefore eighty propulsion events, each resulting in an inclination change of 0.025°. If u_c is 0.01 m/s^2, the time for each maneuver ΔT can be obtained from Eq. (3.56) and is equal to 322 seconds. The total Δv for all eighty maneuvers

is therefore equal to 257.7 m/s. The linear programming approach predicted a Δv of 253.9 m/s, yielding an error of 1%. These examples, when combined with the mathematical validation provided above, bound the accuracy of the linearization approach in predicting the Δv required for a low Earth orbit transfer. The predicted Δv can then be used to determine the reachability of one orbit from another.

3.6 Objectives

DISCO-Tech has several standard objective functions that are commonly used in constellation design problems, such as coverage metrics and cost. Custom user-defined objectives can be developed and included as needed. Some application-specific objectives are described in Chapters 5 and 6.

3.6.1 Coverage

Coverage metrics, also called revisit metrics, quantify the frequency and/or duration of accesses between satellites and ground points. The calculation of coverage metrics begins with the determination of rise and set times for every satellite-ground station combination. The method used for calculating the rise and set times is adapted from the methods developed by Alfano [8]. Alfano uses a coarse time step to find the satellite positions over time, then uses quintic interpolation to approximate the rise-set times of the satellites between the steps of the propagation. They developed equations describing constraints on the maximum range, minimum and maximum elevation angles, and minimum and maximum azimuthal angles. For the purposes of this work, only the maximum range and minimum elevation equations

are used,

$$f_R(t) = R(t) - R_{\text{LIM}} \tag{3.57}$$

$$f_\phi(t) = \left(\cos^{-1}\left\{ \frac{\cos[\phi(t)]}{R(t)} \right\} - \phi(t) \right) - \left\{ \cos^{-1}\left[\frac{\cos(\phi_{\text{LIM}})}{R(t)} \right] - \phi_{\text{LIM}} \right\} \tag{3.58}$$

where $R(t)$ is the range from a satellite to a ground station at time t, $\phi(t)$ is the elevation angle from the ground station to the satellite, R_{LIM} is the maximum allowable range, and ϕ_{LIM} is the minimum allowable elevation angle. Additionally, a sensor with a finite field of view has a maximum off-boresight angle at which the satellite can successfully access a ground point. Therefore, an equation for a constraint on the maximum off-boresight angle of the ground station with respect to the satellite is developed. The off-boresight angle $\theta(t)$ can be calculated as

$$\theta(t) = \cos^{-1}\left[\frac{\mathbf{R}(t) \cdot \mathbf{p}(t)}{\|\mathbf{R}(t)\|} \right] \tag{3.59}$$

where $\mathbf{R}(t)$ is the vector from the satellite to the ground station and $\mathbf{p}(t)$ is the unit vector in the direction of the boresight axis of the sensor.

For the off-boresight angle constraint, the equation is a slight modification of Eq. (3.58),

$$f_\theta(t) = \left(\cos^{-1}\left\{ \frac{\cos[\theta(t)]}{R(t)} \right\} - \theta(t) \right) - \left\{ \cos^{-1}\left[\frac{\cos(\theta_{\text{LIM}})}{R(t)} \right] - \theta_{\text{LIM}} \right\} \tag{3.60}$$

where θ_{LIM} is the maximum allowable off-boresight angle.

To find the intervals that contain rise and set times without interpolating through every time step, the time steps in the coarse propagation for each satellite-station combination are found for which $f_R(t) \leq 0$, $f_\phi(t) \leq 0$, and $f_\theta(t) \geq 0$. During these times, the satellite can access the ground station. Of these time steps, the times are found that are near the beginning interval of access by finding points that are more than one time step away from

the previous point satisfying the constraints. Likewise, the times that are near the end of an interval are found by finding points that are more than one time step away from the next point satisfying the constraints. Either $f_R(t)$, $f_\phi(t)$, or $f_\theta(t)$, whichever is most restrictive at each point, is used for the quintic interpolation as described by Alfano to compute the rise and set times [8]. The rise/set time occurs when the most restrictive function is equal to zero.

For each ground station, a matrix is constructed to describe the access to that station over time. The rows correspond to the sorted rise and set times, while the columns correspond to the satellites. The matrix is binary such that a one in the (i, j) place indicates that the j-th satellite can access the ground station from the i-th time until the $(i+1)$-th time. There is a corresponding vector of times, T_k, to match the rows of the matrix. As an example, the k-th station is accessible by three satellites. The scenario starts at time zero. Satellite 1 can access the station from 5 minutes after the scenario start to 12 minutes after the scenario start. Satellite 2 can access the station from 10 minutes to 20 minutes. Satellite 3 can access the station from 30 minutes to 35 minutes. The scenario ends at 60 minutes. The resulting time vector T_k and access matrix A_k would be

$$T = \begin{Bmatrix} 0 \\ 5 \\ 10 \\ 12 \\ 20 \\ 30 \\ 35 \\ 60 \end{Bmatrix}, \quad A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.61}$$

An access array A is formed by concatenating the accesses for each station. A time matrix T is also constructed. T_{ik} is the i-th rise or set time for the k-th station. A_{ijk} is one if the j-th satellite can access the k-th ground station between T_{ik} and $T_{(i+1)k}$. The length of each time step is calculated as $\Delta T_{ik} = T_{(i+1)k} - T_{ik}$. The access array and time matrix are used to calculate revisit metrics and in the resilience calculations outlined in Chapter 5.

To calculate various revisit metrics, the information in the access array is converted into gap times g_{km} (describing the length of the m-th gap in coverage for the k-th ground station) and a_{kn} (describing the length of the n-th access for the k-th ground station). An access period is only established if a sufficient number of satellites are in view of the ground station simultaneously. The scenario duration is denoted as D. Some common coverage metrics are listed below. The metrics r_k are calculated on a per-ground station basis. The overall value for the mission can be obtained by taking either the average, maximum, or minimum value across all ground stations.

Maximum revisit time: Maximum revisit time is the longest consecutive amount of time a ground station goes without access to the required number of satellites [142],

$$r_k = \max_m g_{km} \tag{3.62}$$

Average revisit time: Average revisit time is the average of the gap times for a ground station [142]. If there are n_{gk} gaps, the average revisit time is

$$r_k = \frac{\sum_{m=1}^{n_{gk}} g_{km}}{n_{gk}} \tag{3.63}$$

Time average gap: The time average gap (TAG) is a weighted average of the gap durations. A related metric is mean time to access (MTTA), which is the average time until next access for

a given ground point when starting from an arbitrary time in the scenario. Mathematically, MTTA is half of TAG. This concept is illustrated in Fig. 3.19. TAG is the area below the gap duration curve, while MTTA is the area under the time until access curve. Since both the base and height of the triangles formed by the time until access curve are equal to the gap duration, the area under this curve is half that of the area under the gap duration curve. The TAG of a ground point is defined as [143]:

$$r_k = \frac{\sum_{m=1}^{n_{gk}} g_{km}^2}{D} \tag{3.64}$$

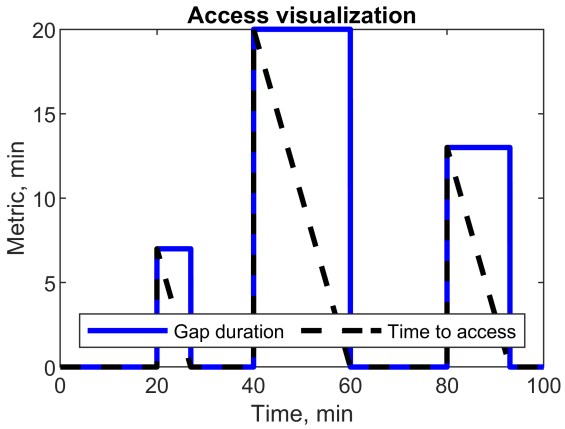


Figure 3.19: Curves for gap duration and time until access.

Average access duration: Average access duration is the average of the access times for a ground station. If n_{ak} is the number of accesses for the k-th ground station, the average access duration is [142]

$$r_k = \frac{\sum_{n=1}^{n_{ak}} a_{kn}}{n_{ak}} \tag{3.65}$$

Percent access: Percent access is the ratio between the amount of time a ground station is

covered and the scenario duration [142],

$$r_k = \frac{\sum_{n=1}^{n_{ak}} a_{kn}}{D} \tag{3.66}$$

Average number of assets: This metric gives the average over time of the number of satellites available to a ground station. It is useful when comparing constellations that provide continuous coverage. If n_{ik} is the number of satellites available to ground station k at time i and $i_{max,k}$ is the number of time steps for ground station k, the average number of assets is [142]

$$r_k = \frac{\sum_{i=1}^{i_{max,k}} n_{ik} \Delta T_{ik}}{D} \tag{3.67}$$

Minimum number of assets: This metric gives the minimum number of satellites available to a ground station at any point in time. The minimum number of assets is [142]

$$r_k = \min_i n_{ik} \tag{3.68}$$

Dilution of precision: The dilution of precision (DOP) metrics do not use the access array. Instead, they are a function of the relative positions of the satellites and the ground stations. DOP is used to evaluate the performance of PNT systems and provides a numerical representation of the sensitivity of the measured position to errors. DOP calculations require simultaneous coverage of at least four satellites. The number of satellites in view of ground station k at time i is denoted as N. The DOP for ground station k at time i requires knowledge of the unit vectors $\hat{\mathbf{r}}_{ijk} = [\hat{x}_{ijk},\ \hat{y}_{ijk},\ \hat{z}_{ijk}]^T = (\mathbf{r}_{ij} - \mathbf{r}_{ik})/\|\mathbf{r}_{ij} - \mathbf{r}_{ik}\|$. $\mathbf{r}_{ij}$ denotes the j-th satellite position at time i in Cartesian coordinates, while $\mathbf{r}_{ik}$ denotes the k-th ground station position at time i in Cartesian coordinates. The C and Q matrices are constructed

[173],

$$C_{ik} = \begin{bmatrix} \hat{x}_{i1k} & \hat{y}_{i1k} & \hat{z}_{i1k} & -1 \\ \hat{x}_{i2k} & \hat{y}_{i2k} & \hat{z}_{i2k} & -1 \\ \hat{x}_{i3k} & \hat{y}_{i3k} & \hat{z}_{i3k} & -1 \\ \vdots & \vdots & \vdots & \vdots \\ \hat{x}_{iNk} & \hat{y}_{iNk} & \hat{z}_{iNk} & -1 \end{bmatrix} \tag{3.69}$$

$$Q_{ik} = (C^T C)^{-1} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} & \sigma_{xt} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} & \sigma_{yt} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 & \sigma_{zt} \\ \sigma_{xy} & \sigma_{yt} & \sigma_{zt} & \sigma_t^2 \end{bmatrix} \tag{3.70}$$

DISCO-Tech deals with three types of DOP. Positional dilution of precision (PDOP) quantifies the effects of the satellite positions on measurement accuracy. Time dilution of precision (TDOP) quantifies the effects of timing on measurement accuracy. Geometric dilution of precision (GDOP) combines the two to represent the impact of both position and timing on accuracy [173].

$$PDOP_{ik} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \tag{3.71}$$

$$TDOP_{ik} = \sigma_t \tag{3.72}$$

$$GDOP_{ik} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_t^2} \tag{3.73}$$

DOP values can then be rolled up over time and space as desired (e.g. minimum, maximum, average).

DOP threshold: A secondary set of metrics using the DOP calculations are the DOP threshold metrics, which measure the percent time a ground station does not possess the required DOP. If the threshold DOP value is DOP_T, the DOP threshold metric is

$$r_k = \frac{1}{D}\int_0^D \delta_k(t)dt \tag{3.74}$$

$$\delta_k(t) = \begin{cases} 1, & DOP_k(t) > DOP_T \\ 0 & DOP_k(t) \leq DOP_T \end{cases} \tag{3.75}$$

3.6.2 Cost

Cost modeling for space systems in early phases of development is imprecise. Few satellite components are commercial off-the-shelf (COTS) parts that have set prices. These COTS parts are normally only available for CubeSats or other small satellites. If exact pricing is available for specific components, that pricing is used in estimating cost. Otherwise, cost models based on historic satellite costs must be used. These cost models are reliant upon the mass estimates for the satellite developed in section 3.4.

Three different cost models are employed by DISCO-Tech. The Small Satellite Cost Model (SSCM) is used for satellites with dry masses of less than 400 kg. [165]. If a satellite's dry mass is greater than 400 kg, the Unmanned Space Vehicle Cost Model (USCM) is used. Both models provide cost estimating relationships (CERs) for the spacecraft bus, integration and testing, programmatic overhead, launch operations support, and ground operations support. USCM provides a CER for communications payloads, while SSCM estimates payload cost as a fraction of bus cost. The NASA Instrument Cost Model (NICM) is combined with the other two models to provide cost estimates for non-communications payloads. NICM provides CERs for optical payloads (both Earth observation and space imaging), microwave

payloads (such as radars, altimeters, scatterometers, and sounders), particle detectors (such as plasma detectors), and field detectors (such as electric and magnetic field detectors) [165]. DISCO-Tech does not enforce hard bounds on the input ranges used by the models, so abnormally large or small systems may require extrapolation of the model outside of its reference data set.

USCM breaks costs down into nonrecurring costs (NRC) and recurring costs (RC), but SSCM and NICM do not. When using SSCM, DISCO-Tech assumes that all costs are recurring except for launch operations support costs, assuming all launches occur simultaneously. NICM wrap costs (testing, development, etc.) are assumed to be nonrecurring, while all other costs are assumed to be recurring.

The total spacecraft cost estimates are then calculated according to the total number of satellites n_s of each type and a learning rate of 0.95 [165],

$$\text{Spacecraft cost} = NRC + RC * n_s^{1+\log_2(0.95)} \tag{3.76}$$

Launch costs are then estimated using a linear programming launch manifestation method as described in Appendix A. A database of current and future launch vehicles and sites are used. The sum of launch costs (LC) and satellite costs is the nominal cost:

$$\text{Nominal cost} = LC + NRC + RC * n_s^{1+\log_2(0.95)} \tag{3.77}$$

Expected cost

Expected cost is the average cost over many constellations where the possibility of total satellite failure exists. The time to satellite failure is estimated by a Weibull distribution,

which has a cumulative distribution function of the form

$$CDF = 1 - e^{-(t/\theta)^{\beta}} \tag{3.78}$$

where θ is the scale parameter and β is the shape parameter. These parameters vary based on satellite size. The parameters for the distribution are shown in Table 3.3, assuming t is the time in years since launch [46].

Table 3.3: Weibull parameters for spacecraft time to failure.

Mass	θ	β
<500 kg	8.93×10^5	0.252
500-2,500 kg	1.82×10^4	0.449
>2,500 kg	273	0.693

A Monte Carlo simulation is used to estimate the expected cost. In each run, a lifetime for each satellite is predicted using the Weibull distribution. If the satellite lifetime is less than the mission lifetime, another satellite is generated with a different lifetime dictated by the Weibull distribution to replace the failed satellite. This process is repeated until each satellite has sufficient units to reach the end of the mission life. Satellite cost is then estimated based on a learning rate of 0.95. Launch costs are calculated more simplistically than in the nominal case due to computation concerns. If new satellites must be launched within three months of each other, the cost of launching those satellites on a single vehicle is calculated if possible. Otherwise, the cheapest launch vehicle for each satellite is selected. The cost is then averaged over all the runs to get the expected cost.

3.7 Summary

Two of the most common DISCO-Tech objectives, coverage and cost, were discussed. DISCO-Tech calculates rise and set times for satellite/ground station combinations by interpolation, then uses the rise and set times to calculate various revisit metrics. The scenarios in Chapters 4 and 5 optimize over one or more revisit metrics. DISCO-Tech uses industry-standard cost models to estimate nominal spacecraft cost, then generates an optimal launch manifest to select launch vehicles and calculate launch costs.. Expected cost can also be calculated based on satellite failure rates. Due to the modular nature of DISCO-Tech, additional objectives can be formulated as needed for specific applications. Chapter 6, for example, uses a custom objective function to calculate sea surface height measurement error.

Chapter 4

Applications: Disaggregation optimization

4.1 Introduction

It is desirable to have an optimization methodology that is applicable for multiple types of disaggregated space system. Existing methodologies can handle one or more types of disaggregation but not all forms of disaggregation. Homogeneous multi-orbit disaggregation can be optimized using genetic algorithms [6], full enumeration of a discrete design space [118], or analytical solutions like Walker constellations. Functional disaggregation across multiple orbits has been analyzed using genetic algorithms to explore predefined architectures [152]. Various levels of fractionation for predefined payloads can be compared using full enumeration of the design space [96], machine learning [67], or evolutionary algorithms [133].

By examining the literature, it is clear that there are gaps in capability. The first gap is that no existing tool has the ability to dynamically optimize payload distribution for multi-orbit, heterogeneous constellations. The second gap is that no existing tool simultaneously optimizes both the payload distribution and the payload parameters. DISCO-Tech seeks to fill these gaps in capability.

The first scenario is a Mars communications and positioning constellation optimization prob-

lem. It optimizes payload and orbital parameters over a predefined set of architectures and serves as a confirmation that genetic algorithms can perform space mission optimization and demonstrates the cost and coverage modules used by DISCO-Tech. The second case is an optimization of a multi-orbit constellation providing coverage over a symmetric latitude band. This case examines the performance of two different methods of varying the number of orbital planes and compares them to a known solution, confirming the ability of a genetic algorithm to optimize various levels of multi-orbit disaggregation. The final case optimizes fractionation of a fixed set of payloads using a variable-length genetic algorithm. The results are compared to a full enumeration of the design space. This problem confirms the ability of a genetic algorithm to optimize payload distribution and varying levels of disaggregation for heterogeneous systems. Together, these test cases confirm the ability of the genetic algorithm used by DISCO-Tech to optimize payload parameters, orbital parameters, and payload distribution simultaneously. This chapter addresses the first research question: How can multiple levels of disaggregation be easily represented and optimized?

4.2 Scenario 4-a: Mars optimization

4.2.1 Purpose and background

To set the foundation for more complicated disaggregated analyses, it is desirable to confirm that genetic algorithms can successfully compare architectures for test problems. A problem with a single objective and finite number of architectures is proposed. This scenario optimizes the design of a Mars-orbiting constellation. The constellation's purpose is to provide baseline communications and position, navigation, and timing (PNT) capabilities to support a theoretical human settlement on Mars. Due to the difficulty of transporting equipment to

the planet's surface, it is likely that much of Mars' infrastructure will be space-based. Optimization was performed to find a constellation that could: 1) provide voice communications for 5,000 simultaneous users, 2) provide position, navigation, and timing (PNT) capabilities globally, and 3) enable data transmission back to Earth with speeds of 1 Mbps. This scenario was originally presented in Ref. 114 and is re-presented here.

4.2.2 Methodology

This scenario uses an earlier version of the DISCO-Tech methodology in which mass modelling is based only on the payload power. The power used by the communications payload is approximated as twice the transmitted power to account for the inefficiency in the amplifier [165]. A relationship between payload power and satellite mass was derived by Springmann and de Weck for nongeosynchronous communications satellites as

$$M_{dry} = 7.5 P_{PL}^{0.65} \tag{4.1}$$

where P_{PL} is the payload power in watts and M_{dry} is the spacecraft dry mass in kg. The same research also shows that the payload mass M_{PL} is approximately $M_{PL} = 0.27 M_{dry}$ [139]. This method will underestimate mass for satellites with a deep space transmitter due to its large antenna mass but provides a basic relationship for this validation case.

It is assumed that the navigation satellites will have sufficient knowledge of their own positions from established ground stations or transmitters. A single objective function was used that penalized solutions that did not meet the standards for coverage and communication but did not reward performance above the minimum requirements.

Four potential satellite types are considered for this analysis. DISCO-Tech's ability to dynamically increase or decrease the number of architectures under consideration was not

leveraged for this problem. The first type, NAV, is a PNT only satellite that transmitted at the low data rate of 50 bps. The second type, COMMS, is a communications satellite that has a variable number of channels available to permit multiple accesses. It is presumed that COMMS could transmit the necessary PNT signal with no additional mass, though a low-gain, wide FOV antenna may be required to achieve fourfold coverage. Also, COMMS could be assigned a deep space transmitter and receiver (DST/R), though it is not required. Both NAV and COMMS are assumed to be Walker Delta constellations. The third type, MR, is a Mars-orbiting satellite hosting a deep space transmitter whose sole function was to transmit data to Earth or to a second relay constellation. It is assigned to a near-equatorial circular plane. A fourth type, SR, is added to consider the possibility of having lower-powered Mars-orbiting satellites transmit to relatively near Sun-orbiting satellites that could then transmit to one another to reach a satellite near Earth.

4.2.3 Optimization formulation

Mission definition

The optimizer sets inclusion variables for each of the four types of satellites. It sets orbital parameters for the satellites as well as the number of satellites. For the navigation and communications satellites, the optimizer sets frequency, transmit power, and field of view. Since COMMS can have an optional deep space payload, an inclusion variable for that payload is present. The decision variables for the optimization are shown in Table 4.1.

Table 4.1: Mars scenario decision variables.

DV	Description	Lower limit	Upper limit	Type
x_1	NAV frequency, GHz	$\{0.2, 0.4, 1.5, 2.5, 5.9, 7.9\}$		Discrete
x_2	NAV satellites per plane	1	20	$\mathbb{Z}$
x_3	NAV number of planes	1	20	$\mathbb{Z}$
x_4	NAV Walker phasing	0	19	$\mathbb{Z}$
x_5	COMMS frequency, GHz	$\{0.2, 0.4, 1.5, 2.5, 5.9, 7.9\}$		Discrete

x_6	COMMS number of channels	1	1,000	$\mathbb{Z}$
x_7	COMMS satellites per plane	1	20	$\mathbb{Z}$
x_8	COMMS number of planes	1	20	$\mathbb{Z}$
x_9	COMMS Walker phasing	0	19	$\mathbb{Z}$
x_{10}	MR number of satellites	1	20	$\mathbb{Z}$
x_{11}	SR number of satellites	1	10	$\mathbb{Z}$
y_1	NAV transmit power, W	0.1	50	$\mathbb{R}$
y_2	NAV half-power beamwidth, °	10	180	$\mathbb{R}$
y_3	NAV semimajor axis, km	3,596	21,396	$\mathbb{R}$
y_4	NAV inclination, °	2	90	$\mathbb{R}$
y_5	NAV inclusion variable	0	1	$\mathbb{R}$
y_6	COMMS transmit power, W	0.1	1,000	$\mathbb{R}$
y_7	COMMS half-power beamwidth, °	10	180	$\mathbb{R}$
y_8	COMMS semimajor axis, km	3,596	21,396	$\mathbb{R}$
y_9	COMMS inclination, °	2	90	$\mathbb{R}$
y_{10}	COMMS DST/R inclusion variable	0	1	$\mathbb{R}$
y_{11}	COMMS inclusion variable	0	1	$\mathbb{R}$
y_{12}	MR transmit power, W	10	500	$\mathbb{R}$
y_{13}	MR semimajor axis, km	3,596	21,396	$\mathbb{R}$
y_{14}	MR inclusion variable	0	1	$\mathbb{R}$
y_{15}	SR transmit power, W	10	500	$\mathbb{R}$
y_{16}	SR semimajor axis, km	1.5×10^8	2.3×10^8	$\mathbb{R}$
y_{17}	SR inclusion variable	0	1	$\mathbb{R}$

Optimizer

The following optimizer parameters are used for the variable-length analysis.

- **Optimizer:** Fixed length genetic algorithm
- **Initial population:** 300
- **Tournament competitors:** 1, with weighted probability of selection. The odds of the n-th most fit candidate being selected for parenthood out of N potential parents are [71]

$$P_n = \frac{N - n + 1}{\sum_{i=1}^{N} i} = \frac{2(N - n + 1)}{N(N + 1)} \tag{4.2}$$

- **Mutation rate:** 5%
- **Length change rate:** 0%
- **Expansion rate:** 0%

- **Nondominated generations for convergence:** 50
- **Number of restarts:** 1
- **Maximum archive size:** 3
- **Epsilon boxes:** N/A
- **Objective limits:** Inf

Dynamics

The following dynamics parameters are used.

- **Propagator:** Fast J4 propagator
- **Central body:** Mars
- **Scenario epoch:** Nov 30, 2017, 00:00
- **Scenario duration:** 10 days
- **Propagation time step:** 86.4 sec
- **Satellite pointing:** Nadir

Constraints

Three constraints are used in this scenario.

Constraint #1, communications channel capacity: The communications satellites must be able to support 5,000 simultaneous users. Furthermore, the satellites must support crosslinking with other communications satellites. The number of simultaneous transmissions achievable by a satellite must be at least equal to the number of expected users for that satellite plus some factor that accounts for the necessary crosslinking between satellites. It is assumed that users are more likely to communicate with users in nearby regions. If the crosslinking factor is f_c, the constraint is

$$g_1(\mathbf{x}) = 5,000 f_c/(x_7 x_8) - x_6 \leq 0 \tag{4.3}$$

Constraint #2–4, link budget constraints: The link budget must be satisfied for voice communications, PNT transmission, and deep space communications. Voice communications require a data rate of 9.6 kbps per channel and a maximum bit error rate of 10^{-3}, which are standard requirements for satellite voice communications used by constellations such as Globalstar [98]. The ground segment of the communications link is modeled after current tactical radios, with a transmit power of 3.2 W and a gain of 1 dBi. PNT requires a data rate of 50 bps and a maximum bit error rate of 10^{-5}. It assumes a ground station gain of -4 dBi. The deep space relay must transmit 1 Mbps across the maximum Earth-Mars distance of 2.5 AU. The deep space signal is assumed to be 8 GHz. The data transmission requirement was checked by examining the multistage link budget connecting a theoretical Mars high-powered ground station to the Deep Space Network on Earth in the worst case scenario when Mars and Earth are separated by 2.5 AU. The uplink from the Mars ground station to the satellite is assumed to transmit at 8 GHz and have up to 100 W of power. The ground antenna is assumed to be a 6 m parabolic antenna with a half power beamwidth (HPBW) of 0.36°. The satellite-based deep space communications unit is also assumed to have a 6 m antenna, with the power set as a variable to be determined during the optimization. The maximum acceptable bit error rate for deep space communications is 10^{-7}. The relay can either go directly from the Mars orbiting satellites to Earth or can go from Mars to the Sun relay satellites to Earth. It is assumed that only one satellite communicates with the deep space network at a time due to limitations on availability, so the deep space data rate is independent of the number of deep space transmitters. A (255, 233) Reed-Solomon coding scheme is assumed for the deep space communications, but no coding is used for the voice communications.

If $BER(\mathbf{x},\mathbf{y}, S, r)$ denotes the bit error rate for the satellites in set S with data rate r, the constraint functions can be adapted from Eq. (3.33) in subsection 3.5.1:

$$g_2(\mathbf{x},\ \mathbf{y}) = BER(\mathbf{x},\mathbf{y}, COMMS, x_6 * 9.6\text{kbps}) - 10^{-3} \leq 0 \tag{4.4}$$

$$g_3(\mathbf{x},\ \mathbf{y}) = BER(\mathbf{x},\mathbf{y}, S_{CN}, 50\text{bps}) - 10^{-5} \leq 0 \tag{4.5}$$

$$g_4(\mathbf{x},\ \mathbf{y}) = BER(\mathbf{x},\mathbf{y}, S_{CMS}, 1\text{Mbps}) - 10^{-7} \leq 0 \tag{4.6}$$

S_{CN} is $\{\text{COMMS}, \text{NAV}\}$ if the solution includes both communications and navigation satellites and $\{\text{COMMS}\}$ if the solution does not include navigation satellites. S_{CMS} is the set of all satellites with deep space relays, which may include COMMS, MR, and SR.

Objective

This scenario has one objective—the sum of cost and revisit metrics. Cost is calculated according to the Unmanned Space Vehicle Cost Model (USCM) and is reported in billions of dollars (Sect. 3.6.2). Launch costs are estimated based on previous launches to Mars. Cost is denoted $C(\mathbf{x},\mathbf{y})$. For the coverage analysis, target points are evenly distributed in latitude an longitude around the planet. The maximum revisit time is calculated for each target point using Eq. (3.62). $R_i(\mathbf{x},\mathbf{y}, S, n)$ is the maximum revisit time in days for the i-th target from the satellites in set S with n-fold coverage. Single-fold revisit time is calculated for the communications satellites. Four-fold revisit time is calculated for the PNT satellites, since four satellites are necessary to provide accurate positioning information. The revisit time is averaged across all targets and scaled by 1,000. If there are n_g target points, the objective function is

$$f_1(\mathbf{x},\mathbf{y}) = C(\mathbf{x},\mathbf{y}) + 1,000\frac{\sum_{i=1}^{n_g} R_i(\mathbf{x},\ \mathbf{y}, \lfloor x_{11} \rceil COMMS, 1)}{n_g} \tag{4.7}$$

$$+ 1,000\frac{\sum_{i=1}^{n_g} R_i(\mathbf{x},\ \mathbf{y}, \{\lfloor x_5 \rceil NAV, \lfloor x_{11} \rceil COMMS\}, 4)}{n_g}$$

The large multipliers on revisit time drive the optimizer toward solutions providing continuous coverage.

Mathematical summary

The mathematical formulation for the Mars scenario is shown below.

Minimize

$$f_1(\mathbf{x},\mathbf{y}) = C(\mathbf{x},\mathbf{y}) + 1,000\frac{\sum_{i=1}^{n_g} R_i(\mathbf{x},\ \mathbf{y}, \lfloor x_{11} \rceil COMMS, 1)}{n_g} + 1,000\frac{\sum_{i=1}^{n_g} R_i(\mathbf{x},\ \mathbf{y}, \{\lfloor x_5 \rceil NAV, \lfloor x_{11} \rceil COMMS\}, 4)}{n_g}$$

such that

$$x_1 \in \{0.2, 0.4, 1.5, 2.5, 5.9, 7.9\} \qquad y_7 \in [10, 180]$$
$$x_2 \in [1, 20] \cap \mathbb{Z} \qquad y_8 \in [3,596, 21,396]$$
$$x_3 \in [1, 20] \cap \mathbb{Z} \qquad y_9 \in [2, 90]$$
$$x_4 \in [0, 19] \cap \mathbb{Z} \qquad y_{10} \in [0, 1]$$
$$x_5 \in \{0.2, 0.4, 1.5, 2.5, 5.9, 7.9\} \qquad y_{11} \in [0, 1]$$
$$x_6 \in [1, 1,000] \cap \mathbb{Z} \qquad y_{12} \in [10, 500]$$
$$x_7 \in [1, 20] \cap \mathbb{Z} \qquad y_{13} \in [3,596, 21,396]$$
$$x_8 \in [1, 20] \cap \mathbb{Z} \qquad y_{14} \in [0, 1]$$
$$x_9 \in [0, 19] \cap \mathbb{Z} \qquad y_{15} \in [10, 500]$$
$$x_{10} \in [1, 20] \cap \mathbb{Z} \qquad y_{16} \in [1.5 \times 10^8, 2.3 \times 10^8]$$

$$x_{11} \in [1, 10] \cap \mathbb{Z} \qquad y_{17} \in [0, 1]$$

$$y_1 \in [0.1, 50] \qquad g_1(\mathbf{x}) = 5,000 f_c/(x_7 x_8) - x_6 \leq 0$$

$$y_2 \in [10, 180] \qquad g_2(\mathbf{x},\ \mathbf{y}) = BER(\mathbf{x},\mathbf{y}, COMMS, x_6 * 9.6\text{kbps})$$

$$y_3 \in [3,596, 21,396] \qquad - 10^{-3} \leq 0$$

$$y_4 \in [2, 90] \qquad g_3(\mathbf{x},\ \mathbf{y}) = BER(\mathbf{x},\mathbf{y}, S_{CN}, 50\text{bps})$$

$$y_5 \in [0, 1] \qquad - 10^{-5} \leq 0$$

$$y_6 \in [0.1, 1,000] \qquad g_4(\mathbf{x},\ \mathbf{y}) = BER(\mathbf{x},\mathbf{y}, S_{CMS}, 1\text{Mbps}) - 10^{-7} \leq 0$$

4.2.4 Results and conclusions

The simulation determined that the most cost effective solution was to use a single type of satellite, a communications satellite augmented with an additional deep space transmitter and receiver. The resulting satellite and constellation had the parameters shown in Table 4.2, with visualization shown in Fig. 4.1. The satellite will have a mass of approximately 120 kg, though the mass estimate does not include specific antenna mass modeling and therefore underestimates the mass of the deep space communicator. The estimated costs for the constellation, in 2020 dollars (FY2020), are $1.5 billion for development, $32 million for the first flight unit, $1.1 billion for all satellites, and $266 million for fifteen PSLVs, giving a total mission cost of $2.9 billion.. The high development cost was driven by the number of channels and precluded the use of multiple types of satellites.

Table 4.2: Mars solution properties

Properties	**Value**
Comms frequency (GHz)	0.2
Comms transmit power (W)	6.5
Comms HPBW (deg)	26
Comms number of channels	173
Semimajor axis (km)	9,818
Inclination (deg)	85.2
Satellites per plane	9
Number of planes	5
Walker phasing parameter	1
Deep space transmit power (W)	29.3

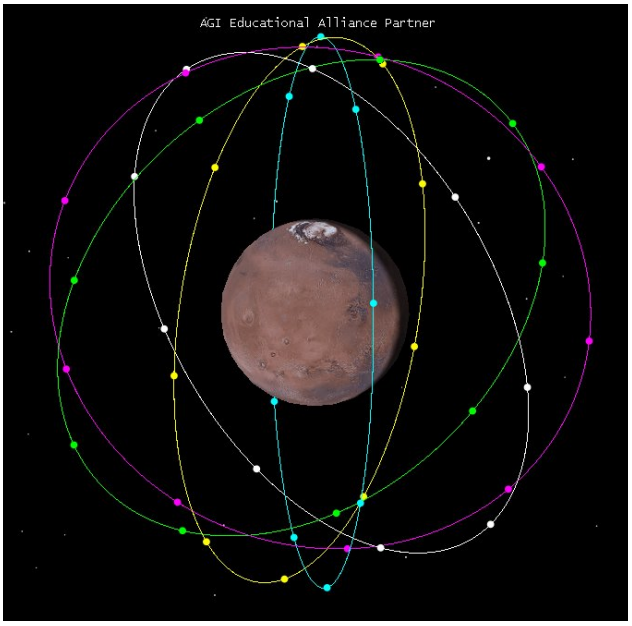


Figure 4.1: Visualization of final constellation.

The constellation provides continuous four-fold coverage over the entire planet. The maximum geometric dilution of precision of the constellation ranges between 1.2 and 2.1 for various latitudes. The voice communications have an uplink bit error rate of $6.0 * 10^{-5}$ and a downlink bit error rate of $6.7 * 10^{-11}$. The bit error rate for the PNT data is insignificant due to the low data rate. The bit error rate for data transmission to Earth is $4.7 * 10^{-8}$. The deep space communications transmitter had a power of 29.3 W or 14.7 dBW. The antenna

gain was 52.6 dBi for an Effective Isotropic Radiated Power (EIRP) of 67.3 dBW without line loss. This result compares favorably with the Mars Reconnaissance Orbiter, which has a transmit power of 100 W (20 dBW) and an antenna gain of 48 dBi for an EIRP of 68 dBW [148].

The link budgets for the satellite are shown in Tables 4.3–4.5.

Table 4.3: Solution voice downlink budget.

Parameter	**dB Value**	**Supporting Value**
Frequency		200 MHz
Transmit power	8.13 dBW	6.5 W
Peak transmit antenna gain	16.34 dBi	
Transmit FOV loss	-2.78 dB	
Transmit line loss	-0.75 dB	
EIRP	20.94 dBW	
Number of channels		173
EIRP per channel	**-1.44 dBW**	
Space loss	-155.36 dB	
Atmospheric loss	-0 dB	
Implementation loss	-1.2 dB	
Polarization loss	-0.75 dB	
Losses	**-157.31 dB**	
Receiver gain	**1.00 dBi**	
Boltzmann's constant, k		1.38×10^{-23} J/K
1/k	228.60 dBK/J	7.2×10^{22} K/J
Noise temperature, T		231.15 K
1/T	-23.64 dB/K	4.32×10^{-3} 1/K
Data rate per channel, R		9.6 kbps
1/R	-39.82 dB/Hz	1.04×10^{-4} s/bit
1/(kRT)	**165.14 dB/Hz/J**	
E_b/N_0	**7.39 dB**	
Required E_b/N_0	**4.78 dB**	
Link margin	**2.62 dB**	

Ultimately, a single satellite type was the most viable due to development and launch costs, which could have eliminated the two relay satellite types. The number of communications satellites needed made the additional PNT satellites unnecessary. If a satellite were made with commercial off-the-shelf components and the cost modeled accordingly, it is possible that a multiple satellite solution would become preferable. Furthermore, if the target area of the communications constellation were reduced to the equatorial region, it is likely that

Table 4.4: Solution voice uplink budget.

Parameter	**dB Value**	**Supporting Value**
Frequency		200 MHz
Transmit power	5.05 dBW	3.20 W
Peak transmit antenna gain	1 dBi	
Transmit FOV loss	-3 dB	
Transmit line loss	-0.75 dB	
EIRP	2.30 dBW	
Number of channels		1
EIRP per channel	**2.30 dBW**	
Space loss	-155.36 dB	
Atmospheric loss	-0 dB	
Implementation loss	-1.2 dB	
Polarization loss	-0.75 dB	
Losses	**-157.31 dB**	
Receiver gain	**13.56 dBi**	
Boltzmann's constant, k		1.38×10^{-23} J/K
1/k	228.60 dBK/J	7.2×10^{22} K/J
Noise temperature, T		468.75
1/T	-26.71 dB/K	2.13×10^{-3}
Data rate per channel, R		9.6 kbps
1/R	-39.82 dB/Hz	1.04×10^{-4} s/bit
1/(kRT)	**162.07 dB/Hz/J**	
$\mathbf{E_b/N_0}$	**20.63 dB**	
Required $\mathbf{E_b/N_0}$	**4.78 dB**	
Link margin	**15.85 dB**	

the lower cost PNT satellites would be a superior choice to provide the remaining coverage. Regardless, this scenario demonstrated the capability of DISCO-Tech to compare the performance of heterogeneous and homogeneous constellations. The fact that the optimal solution was a homogeneous constellation highlights the need to avoid prematurely eliminating architecture options due to a preconceived notion of what the "best" solution may be. Additionally, the ability to simultaneously optimize orbital parameters and payload parameters was demonstrated.

Sensitivity to mass model

The initial analysis used a mass model that was developed based on historical data for nongeosynchronous communications satellites. The mass was driven by the payload power,

Table 4.5: Solution data downlink budget.

Parameter	**dB Value**	**Supporting Value**
Frequency		8 GHz
Transmit power	14.67 dBW	29.3 W
Peak transmit antenna gain	53.48 dBi	
Transmit FOV loss	-0.84 dB	
Transmit line loss	-0.75 dB	
EIRP	66.56 dBW	
Number of channels		1
EIRP per channel	**66.56 dBW**	
Space loss	-281.97 dB	
Atmospheric loss	-0.24 dB	
Implementation loss	-1.2 dB	
Polarization loss	0 dB	
Losses	**-283.41 dB**	
Receiver gain	**69.68 dBi**	
Boltzmann's constant, k		1.38×10^{-23} J/K
1/k	228.60 dBK/J	7.2×10^{22} K/J
Noise temperature, T		35.00 K
1/T	-15.44 K	2.86×10^{-2}
Data rate per channel, R		1 Mbps
1/R	-60 dB/Hz	1×10^{-6} s/bit
1/(kRT)	**153.16 dB/Hz/J**	
E_b/N_0	**5.99 dB**	
Required E_b/N_0	**5.95 dB**	
Link margin	**0.05 dB**	

which was in turn driven by the transmit power. Because the antenna aperture was not used in estimating the payload mass, the mass for the deep space communications payload is underestimated.

New mass and cost estimates are generated using the standard DISCO-Tech sizing methodology for communications satellites described in section 3.4. The estimated satellite dry mass is 340 kg, and the estimated satellite wet mass is 625 kg. Because the satellite dry mass is less than 400 kg, SSCM is used to estimate cost. The estimated cost is \$3.2B (FY2020). SSCM does not distinguish between recurring and nonrecurring costs. The estimated cost for the first flight unit is \$81M. It is no longer possible to use PSLVs due to the increased spacecraft mass. Instead, the Falcon Heavy was determined to be the most cost effective option. Five Falcon Heavy rockets are required at \$93M per rocket.

If USCM is used instead, the estimated cost is $4.3B (FY2020). USCM estimates the nonrecurring development costs to be $1.8B. The recurring cost for the first flight unit is $60M, and the cost for all satellites is $2B. Both models estimate higher costs than the $2.9B predicted in the original analysis due to the higher satellite mass estimates.

The cost model may affect the disaggregation profile chosen by the optimizer. Because USCM predicts high development costs, it is less likely to produce a heterogeneous constellation. A constellation of 45 communictions-only satellites and 2 deep space-only satellites has a USCM-predicted cost of $4.8B. By contrast, since SSCM is based only on mass and does not separately calculate development costs, it is likely to give preference to a heterogeneous constellation. A constellation of 45 communictions-only satellites and 2 deep space-only satellites has a SSCM-predicted cost of $1.5B due to the decrease in average mass when the deep space payload is included on only some of the satellites.

These numbers does not include an assessment of the benefit of redundant systems. In the initial analysis, all deep space communications units were required to have a fixed antenna size of six meters and were required to be capable of transmitting 1 Mbps to Earth at the maximal Earth-Mars separation. If a larger number of deep space communications payloads are present, it may be possible to reduce the data rate of each individual payload if the Deep Space Network can communicate with multiple satellites at a time. Decreasing the data rate would allow either a decrease in transmit power or a decrease in antenna size, both of which would reduce the mass and therefore the predicted cost. Assuming that a constellation with 45 deep space transmitters has half (22) of the constellation transmitting simultaneously, each satellite must then transmit only 45.5 kbps of data. Neglecting any losses due to interference, the decrease in data rate is equivalent to a 13.4 dB increase in the link budget. At this data rate, the antenna diameter can be decreased from 6 m to 1.26 m, drastically reducing the mass of the deep space communications unit. For USCM, this reduces the estimated cost

to $3.3B, cementing the result that the homogeneous constellation is most appropriate. For SSCM, the estimated cost is $2B. The decreased difference between the homogeneous and heterogeneous constellation costs estimated by SSCM may mean that the additional cost of the homogeneous solution is worthwhile to reduce manufacturing complexity and to avoid cost overrun due to additional development.

Sensitivity to communications parameters

The initial analysis made assumptions about communications characteristics when developing a link budget. Transmitter line loss was assumed to be 0.75 dB. No power backoff was included. Loss due to misaligned polarization was assumed to be 0.75 dB. The receiver noise temperature was estimated to be 70 K, and galactic noise and cosmic background radiation were considered. No error-correcting coding was used for the voice communications.

For a more robust downlink voice communications budget, a power backoff of 3 dB is added for the voice communications system. Additionally, a polarization loss of 7 dB is considered to account for misalignment of the polarization of the transmit and receive antennas. A (255, 223) Reed-Solomon code is assumed. The uplink budget is also assumed to include noise due to radiation from the surface of Mars at 293 K. Under these assumptions, the downlink budget is deficient by 6.79 dB due to intial margin in the link budget and the coding gain. To meet the 10^{-3} bit error rate requirement, the transmit power must be increased by a factor of about 5 to 32 W.

The uplink budget has sufficient margin such that it remains unchanged. The deep space link budget does not require power backoff because there is no need for multiple accesses. It also assumed that the deep space transmitter is right-hand circularly polarized, like the Mars Reconaissance Orbiter antenna. Because this polarization matches that of the Deep

Space Network and because the atmospheric effects on polarization at X band are minimal, no polarization loss is assumed.

Due to the increase in transmit power, both the communications subsystem and the power subsystem require increased mass. The dry mass of the satellite becomes 359 kg. USCM estimates a total mission cost of $4.4B, while SSCM estimates a total mission cost of $3.6B. These estimates are less than 10% more than the cost predictions developed for the original link budget using the detailed mass model, so it is unlikely that adjustments in the link budget will affect the disaggregation scheme.

4.2.5 Conclusions

DISCO-Tech produced a homogeneous constellation of 45 satellites that provide continuous communications and positioning capabilities for a settlement on Mars. The satellites also relay scientific data to Earth-based ground stations. The total mission cost was estimated to be $2.9 billion. This scenario demonstrated DISCO-Tech's ability to optimize a constrained space mission with multiple candidate spacecraft types. It also showed DISCO-Tech's ability to estimate cost and coverage performance.

4.3 Scenario 4-b: Symmetric coverage

4.3.1 Purpose and background

When optimizing multiple planes of satellites using genetic algorithms, two strategies for varying the number of planes are the use of variable length genomes and the use of inclusion variables, or flags. This scenario examines the ability of both methods to converge upon

an expected solution. The Walker delta constellation, described in Chapter 2, is known to provide good coverage for global equatorially symmetric target regions [159]. Therefore, the DISCO-Tech methodology's ability to generate a Walker-like constellation is tested. The number of orbital planes is varied using two methods: a variable-length genetic algorithm and a flagging method in which some decision variable sets the inclusion of each plane described by the genome. Additionally, all Walker constellation configurations for a fixed altitude and inclination with 100 or fewer satellites are assessed.

4.3.2 Methodology

The standard DISCO-Tech methodology is used for this scenario. A flagging method is implemented by including a continuous decision variable that takes values between zero and one for each segment. If the value of the flag in a particular solution is less than 0.5, that segment is not considered in the analysis.

4.3.3 Optimization formulation

Mission definition

A satellite constellation with fixed semimajor axis, eccentricity, and inclination is defined. The orbits are assumed to be circular with an altitude of 1,500 km and an inclination of 60°. The conical field of view of the satellite is assumed to be centered at nadir and extend to 50° on either side of nadir. Up to ten planes of satellites are allowed with up to twenty satellites per plane. The satellites are evenly distributed in true anomaly about the orbit, with the true anomaly of one satellite per plane being set by the optimization. The number of satellites, RAAN, and true anomaly are set separately for each orbit.

The decision variables for the variable-length optimization are shown in Table 4.6. The decision variables for the the flagged optimization are shown in Table 4.7.

Table 4.6: Variable-length coverage analysis decision variables.

DV	**Description**	**Lower limit**	**Upper limit**	**Type**
x_{i1}	Number of satellites in orbit	1	20	$\mathbb{Z}$
y_{i1}	Satellite RAAN, °	0	360	$\mathbb{R}$
y_{i2}	Reference true anomaly, °	0	360	$\mathbb{R}$
$i_{\max}$	Number of orbital planes	1	10	$\mathbb{Z}$
for i	plane ID	1	$i_{\max}$	$\mathbb{Z}$

Table 4.7: Flagged coverage analysis decision variables.

DV	Description	Lower limit	Upper limit	Type
x_{i1}	Number of satellites in orbit	1	20	$\mathbb{Z}$
y_{i1}	Satellite RAAN, °	0	360	$\mathbb{R}$
y_{i2}	Reference true anomaly, °	0	360	$\mathbb{R}$
y_{i3}	Inclusion flag	0	1	$\mathbb{R}$
for i	plane ID	1	10	$\mathbb{Z}$

Optimizer

The following optimizer parameters are used for the variable-length analysis.

- **Optimizer:** Variable-length genetic algorithm with epsilon boxing
- **Initial population:** 100
- **Tournament competitors:** 2
- **Mutation rate:** 5%
- **Length change rate:** 10%
- **Expansion rate:** 0%
- **Nondominated generations for convergence:** 20
- **Number of restarts:** 10
- **Maximum archive size:** 10,000
- **Epsilon boxes:** [0.01, 0.01, 1]

- **Objective limits:** [Inf, Inf, Inf]

The flagged analysis has identical parameters except for the length change rate, which is equal to zero, transforming the optimizer into a fixed-length genetic algorithm with epsilon boxing.

Dynamics

The following dynamics parameters are used.

- **Propagator:** Fast J4 propagator
- **Central body:** Earth
- **Scenario epoch:** Oct 9, 2019, 00:00
- **Scenario duration:** 0.5 days
- **Propagation time step:** 86.4 sec
- **Satellite pointing:** Nadir

Constraints

No constraints are used in this scenario.

Objectives

This scenario has three objectives.

Objective #1–2, maximize average and minimum percent access: Coverage is calculated over 200 points evenly spaced in the latitude band between 60°S and 60°N. The percent access over time is calculated for each ground point using Eq. (3.66). The average

of these values is used as an objective, as is the minimum of these values. To maximize these objectives, a multiplier of -1 is included in the objective function. A solution providing continuous coverage over the entire target region would have values of -1 for both objectives. An epsilon value of 0.01 (1%) is used for both objectives.

If Δt_j is the length of time step j, δ_{kj} is a binary variable equal to one if target k can access at least one satellite at time step j and zero otherwise, n_t is the number of time steps, and n_g is the number of target points, the objective functions are

$$f_1(\mathbf{x},\, \mathbf{y}) = -1 \sum_{k=1}^{n_g} \left(\frac{\sum_{j=1}^{n_t} \delta_{kj}(\mathbf{x},\, \mathbf{y}) \Delta t_j}{n_g \sum_{j=1}^{n_t} \Delta t_j} \right) \tag{4.8}$$

$$f_2(\mathbf{x},\, \mathbf{y}) = -1 \min_k \left(\frac{\sum_{j=1}^{n_t} \delta_{kj}(\mathbf{x},\, \mathbf{y}) \Delta t_j}{\sum_{j=1}^{n_t} \Delta t_j} \right) \tag{4.9}$$

Objective #3, minimize number of satellites: This objective is set to explore the design space and examine the performance of constellations of various sizes. The epsilon box for this objective is set at one satellite. The objective function differs slightly between the two scenarios. For the variable-length analysis, it is

$$f_3(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i1} \tag{4.10}$$

For the flagged analysis, the objective is

$$f_3(\mathbf{x},\, \mathbf{y}) = \sum_{i=1}^{10} x_{i1} \lfloor y_{i3} \rceil \tag{4.11}$$

Mathematical summary

The mathematical formulation for the variable-length optimization is shown below.

Minimize

$$f_1(\mathbf{x},\,\mathbf{y}) = -1 \sum_{k=1}^{n_g} \left(\frac{\sum_{j=1}^{n_t} \delta_{kj}(\mathbf{x},\,\mathbf{y}) \Delta t_j}{n_g \sum_{j=1}^{n_t} \Delta t_j} \right)$$
$$f_2(\mathbf{x},\,\mathbf{y}) = -1 \min_k \left(\frac{\sum_{j=1}^{n_t} \delta_{kj}(\mathbf{x},\,\mathbf{y}) \Delta t_j}{\sum_{j=1}^{n_t} \Delta t_j} \right)$$
$$f_3(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i1}$$

such that

$$i_{\max} \in [1, 10] \cap \mathbb{Z}$$
$$x_{i1} \in [1, 20] \cap \mathbb{Z}\ \forall i = 1, 2, \ldots, i_{\max}$$
$$y_{i1} \in [0, 360]\ \forall i = 1, 2, \ldots, i_{\max}$$
$$y_{i2} \in [0, 360]\ \forall i = 1, 2, \ldots, i_{\max}$$

The mathematical formulation for the flagged optimization is shown below.

Minimize

$$f_1(\mathbf{x},\,\mathbf{y}) = -1 \sum_{k=1}^{n_g} \left(\frac{\sum_{j=1}^{n_t} \delta_{kj}(\mathbf{x},\,\mathbf{y}) \Delta t_j}{n_g \sum_{j=1}^{n_t} \Delta t_j} \right)$$
$$f_2(\mathbf{x},\,\mathbf{y}) = -1 \min_k \left(\frac{\sum_{j=1}^{n_t} \delta_{kj}(\mathbf{x},\,\mathbf{y}) \Delta t_j}{\sum_{j=1}^{n_t} \Delta t_j} \right)$$
$$f_3(\mathbf{x},\,\mathbf{y}) = \sum_{i=1}^{10} x_{i1} \lfloor y_{i3} \rceil$$

such that

$$x_{i1} \in [1, 20] \cap \mathbb{Z} \ \forall i = 1, 2, \ldots, 10$$

$$y_{i1} \in [0, 360] \ \forall i = 1, 2, \ldots, 10$$

$$y_{i2} \in [0, 360] \ \forall i = 1, 2, \ldots, 10$$

$$y_{i3} \in [0, 1] \ \forall i = 1, 2, \ldots, 10$$

4.3.4 Results

The nondominated fronts for the variable-length analysis, the flagged analysis, and the exhausive search of the Walker constellation space are shown in Fig. 4.2. The figure displays only the minimum percent access for simplicity's sake. The points that are nondominated (ND) in number of satellites and minimum percent access for a given algorithm are marked with a circle, square, or triangle, while points that are dominated (D) in number of satellites and minimum percent access (but nondominated across all three objectives) are marked with an x. The variable-length GA and flagged GA produced similar fronts, with the variable-length front being slightly better. The variable-length GA took 1,988 generations and 556,277 function evaluations, producing 80 nondominated solutions. The flagged GA took 1,811 generations and 458,315 function evaluations, producing 75 nondominated solutions. The optimization routines performed comparably to the Walker constellations when the number of satellites was less than 20. They failed to perform as well with constellations of medium size, between 20 and 40 satellites. The optimization routines outperformed the Walker constellations in the cases with more than 40 satellites. This performance difference may be explained by the fact that the reference true anomaly was not optimized in the Walker cases. Because the time step for the dynamics is approximately 86 seconds, gaps in coverage with

duration less than the dynamics time step may be missed in the analysis. It is possible that the optimization cases picked reference anomalies that effectively hid some small gaps.

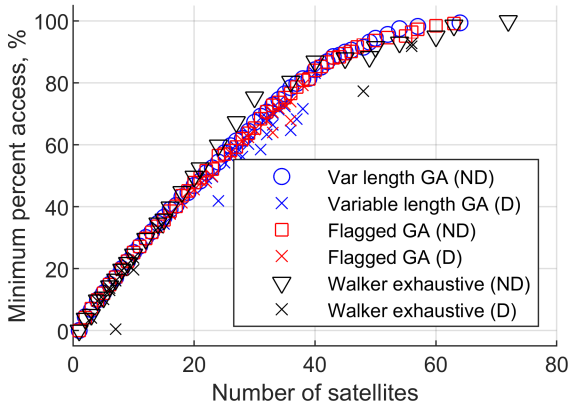


Figure 4.2: Nondominated fronts for convergence test.

The nondominated fronts produced by the two different optimization routines can be compared using several different metrics [86]. Ref. 73 suggested the use of the hypervolume indicator (S-metric) for comparing variable-length and flagged algorithms. Hypervolume measures the dominated space as defined by the front and some reference point. In this case, the reference point is the worst value for each objective found across any generation in either analysis (disregarding flagged cases in which no satellites are set). The objectives are scaled by dividing the objective value by the epsilon value corresponding to that objective. All three objectives are used in the calculation of the hypervolume. The process for calculating hypervolume is described in Ref. 57. For two hypervolumes calculated using the same reference value, a larger hypervolume indicates a better front. The hypervolume for both optimization routines is calculated every ten generations for the first hundred generations and every fifty generations thereafter. Hypervolume is plotted against the total number of function evaluations so far in Fig. 4.3. The hypervolume for the variable-length method is either greater than or approximately equal to the hypervolume for the flagged method. This result suggests that the variable-length method is superior according to the hypervolume

indicator.

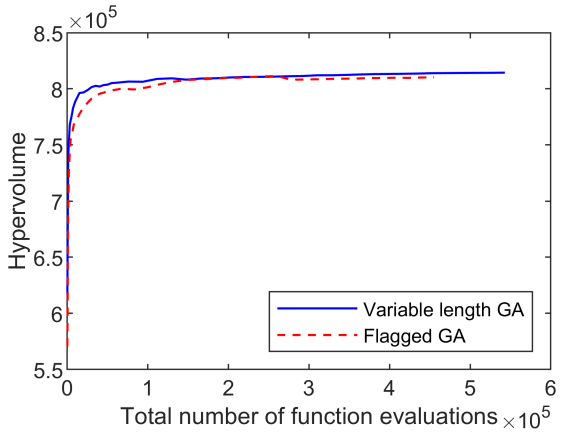


Figure 4.3: Hypervolume indicators for convergence test.

The variable-length analysis produced a solution of 64 satellites that has an average percent access of 99.8% and a minimum percent access of 99.3%. Since the epsilon values for the access metrics are 1%, this solution is in the best possible epsilon box for access. The solution is shown in Fig. 4.4. The 64 satellites are spread across nine planes. There are two planes of one satellite, one plane of eight satellites, and six planes of nine satellites. One of the single satellite planes is very close to a nine satellite plane. The two planes will be considered to be a single plane in the remaining analysis, producing an eight plane constellation. An eight plane Walker constellation has planes spaced every 45 degrees in RAAN. The average spacing of the optimized constellation is 45 degrees (as is the average RAAN spacing of any eight plane constellation), and the standard deviation is 13.5 degrees. The planes with one and eight satellites are anomalous, but because the constellation achieved near-continuous coverage, the solution is not driven to improve further. Because the number of satellites per plane is constant with some exceptions (the one and eight satellite planes) and the spacing has a low standard deviation, the constellation can be considered Walker-like. The fact that the optimization produced a front similar to that of the exhaustive Walker analysis indicates that the variable-length optimization strategy performs well for optimizing constellations

that are disaggregated across multiple planes.

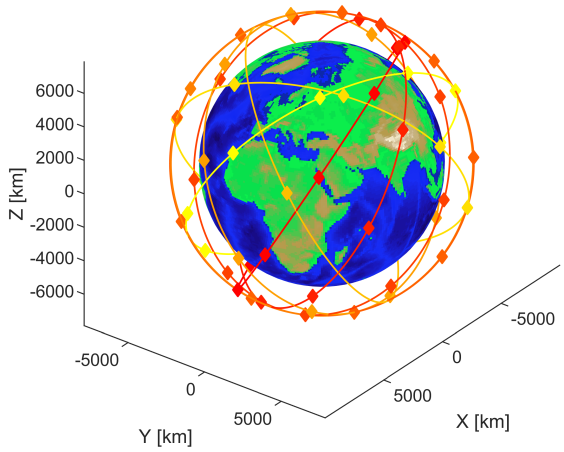


Figure 4.4: Near-continuous coverage constellation from variable-length optimization.

4.3.5 Conclusions

These results show that the variable-length optimization technique outperforms the flagged optimization technique for the case of optimizing relative placement of multiple planes of satellites. The front produced by both optimization techniques is similar by that produced using traditional Walker constellation design strategies. This scenario demonstrates the ability of a variable-length GA to easily represent and optimize multiple levels of disaggregation for multiplane distributed space systems and validates the use of variable-length GAs in the DISCO-Tech methodology.

4.4 Scenario 4-c: Fractionation problem

4.4.1 Purpose and background

The previous example explored the performance of DISCO-Tech with regards to optimization of multi-orbit disaggregated architectures. An additional aspect of DISCO-Tech is its ability to optimize heterogeneous architectures by varying the number of satellites and satellite configurations in an architecture. This form of disaggregation can be combined with multi-orbit disaggregation to form heterogeneous constellations or used with nearby groups of modules or satellites through fractionation or functional disaggregation. This analysis uses a rules-based scenario to assess DISCO-Tech's ability to optimize payload assignments for heterogeneous clusters of modules. The optimization results are compared to the results of an exhaustive search performed by evaluating all combinations of payloads across up to five modules.

4.4.2 Methodology

The DISCO-Tech methodology is used to model and solve the optimization problem. No dynamics simulation or mass estimation are performed. An application-specific cost model is used. Performance and resilience are assessed based on predefined rules. There are five payloads that must be present in the cluster.

Cost model

The cost model used in this analysis assumes fixed payload costs plus scaling factors based on vehicle complexity. Cost is treated as unitless. Each of the five payloads has a base cost.

The base costs are 400 for Payload 1, 500 for Payload 2, 500 for Payload 3, 300 for Payload 4, and 100 for Payload 1. The base cost for the j-th payload is denoted $c_{p,j}$. The cost of the i-th module is

$$c_{s,i} = \alpha \left(\sum_{j=1}^{5} c_{p,j} \delta_{ij} \right)^{\eta} + \beta \tag{4.12}$$

δ_{ij} is one if the j-th payload is present on the i-th module and zero otherwise. α is used to scale costs based on the payload cost and encompasses costs due to bus hardware. η is used to account for resource sharing between sensors. β represents fixed costs. In this analysis, $\alpha = 1.5$, $\eta = 0.9$, and $\beta = 200$.

If n_s is the number of modules, the total launch costs for all the modules are

$$c_L = \gamma + \delta n_s + \epsilon \sum_{i=1}^{n_s} c_{s,i} \tag{4.13}$$

γ includes fixed launch costs like launch support. δ represents launch costs that scale with the number of modules. ϵ represents launch costs that scale with the the cost/mass of the modules. In this analysis, $\gamma = 200$, $\delta = 700$, and $\epsilon = 0.1$. The total cost is the sum of all module costs and the launch costs.

Rules-based performance model

A rules-based performance model is used in this analysis. Performance is assessed on a scale of 0 to 100, with 100 being the best.

1. 10 points each if Payload 1, Payload 2, and Payload 3 are present at least once (30 points)

2. 16 points if Payload 4 is present at least once (16 points)

3. 20 points if Payload 5 is present at least once (20 points)
4. 10 points if both Payload 3 and Payload 4 are present at least once (10 points)
5. 8 points if Payload 1 and Payload 2 are on the same module (8 points)
6. 10 points if either Payload 3 or Payload 5 is present at least once (10 points)
7. 6 points if there are at least two instances of Payload 1, Payload 3, and Payload 5, with duplicates allowed (6 points)

A monolithic satellite would have a score of 100. The condition that Payload 1 and Payload 2 be on the same module means that a fractionated solution can have less than a perfect score even if all payloads are included. Nominal performance is not used as an objective, but performance after the loss of one or two modules is assessed.

4.4.3 Optimization formulation

Mission definition

The optimizer sets inclusion variables for each of the five payloads on each module. A payload is present on the module if the value of the corresponding decision variable is greater than or equal to 0.5. There can be up to five modules.

The decision variables for the fractionation optimization are shown in Table 4.8.

Table 4.8: Rules-based fractionation problem decision variables.

DV	Description	Lower limit	Upper limit	Type
y_{i1}	Payload 1 inclusion variable	0	1	$\mathbb{R}$
y_{i2}	Payload 2 inclusion variable	0	1	$\mathbb{R}$
y_{i3}	Payload 3 inclusion variable	0	1	$\mathbb{R}$
y_{i4}	Payload 4 inclusion variable	0	1	$\mathbb{R}$
y_{i5}	Payload 5 inclusion variable	0	1	$\mathbb{R}$
$i_{\max}$	Number of modules	1	5	$\mathbb{Z}$
for i	module ID	1	$i_{\max}$	$\mathbb{Z}$

Optimizer

The following optimizer parameters are used.

- **Optimizer:** Variable-length genetic algorithm with epsilon boxing
- **Initial population:** 200
- **Tournament competitors:** 4
- **Mutation rate:** 15%
- **Length change rate:** 10%
- **Expansion rate:** 5%
- **Nondominated generations for convergence:** 15
- **Number of restarts:** 10
- **Maximum archive size:** 10,000
- **Epsilon boxes:** [50, 0.2, 0.2, 0.2]
- **Objective limits:** [Inf, Inf, Inf, Inf]

Dynamics

Because this scenario is a payload assignment problem, no dynamics are used and propagation is unnecessary.

Constraints

One constraint is used in this scenario.

Constraint #1, all payloads included: Every payload must be included at least once.

Adapted from Eq. (3.37), the constraint equation is

$$g_{1j}(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{ij} \rceil \leq 0 \forall j \in 1, 2, 3, 4, 5 \tag{4.14}$$

This constraint is repairable. If the constraint is violated for the j-th payload, a random integer $1 \leq i \leq i_{\max}$ is selected. y_{ij} is set to a random value between 0.5 and 1 to indicate that the payload is assigned.

Objectives

This scenario has four objectives.

Objective #1, minimize cost: From Eqs. (4.12) and (4.13), the total cost is

$$f_1(\mathbf{y}) = 200 + 920 i_{\max} + 1.1 \sum_{i=1}^{i_{\max}} 1.65 \left(\sum_{j=1}^{5} c_{p,j} \lfloor y_{ij} \rceil \right)^{0.9} \tag{4.15}$$

The epsilon value for this objective is 50.

Objective #2–4, maximize degraded performance: $P(\mathbf{y}, \mathbf{z})$ is the function describing the performance of the cluster defined by $\mathbf{y}$ when the modules with indices in $\mathbf{z}$ have been removed. Performance is calculated as described in subsubsection 4.4.2. The second objective is to maximize the worst-case performance when a single module is lost. The third objective is to maximize the average performance when a single module is lost. The fourth objective is to maximize the average performance when two modules are lost. The objective functions are multiplied by -1 to transform them into minimization problems. If k and q are integers, the objective functions are

$$f_2(\mathbf{y}) = -1 \min_{k \in [1, i_{\max}] \cap \mathbb{Z}} P(\mathbf{y}, [k]) \tag{4.16}$$

$$f_3(\mathbf{y}) = \frac{-1}{i_{\max}} \sum_{k=1}^{i_{\max}} P(\mathbf{y}, [k]) \tag{4.17}$$

$$f_4(\mathbf{y}) = \frac{-1}{C_2^{i_{\max}}} \sum_{q=k+1}^{i_{\max}} \sum_{k=1}^{i_{\max}} P(\mathbf{y}, [k, q]) \tag{4.18}$$

$C_2^{i_{\max}}$ is the number of combinations of two removals that can be formed from $i_{\max}$ modules. The epsilon value for each objective is 0.2.

Mathematical summary

The mathematical formulation for the fractionation problem is shown below.

Minimize

$$f_1(\mathbf{y}) = 200 + 920 i_{\max} + 1.1 \sum_{i=1}^{i_{\max}} 1.65 \left(\sum_{j=1}^{5} c_{p,j} \lfloor y_{ij} \rceil \right)^{0.9}$$

$$f_2(\mathbf{y}) = -1 \min_{k \in [1, i_{\max}] \cap \mathbb{Z}} P(\mathbf{y}, [k])$$

$$f_3(\mathbf{y}) = \frac{-1}{i_{\max}} \sum_{k=1}^{i_{\max}} P(\mathbf{y}, [k])$$

$$f_4(\mathbf{y}) = \frac{-1}{C_2^{i_{\max}}} \sum_{q=k+1}^{i_{\max}} \sum_{k=1}^{i_{\max}} P(\mathbf{y}, [k, q])$$

such that

$$i_{\max} \in [1, 5] \cap \mathbb{Z}$$

$$y_{i1} \in [0, 1] \ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i2} \in [0, 1] \ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i3} \in [0,1] \ \forall i = 1,2,\ldots,i_{\max}$$

$$y_{i4} \in [0,1] \ \forall i = 1,2,\ldots,i_{\max}$$

$$y_{i5} \in [0,1] \ \forall i = 1,2,\ldots,i_{\max}$$

$$g_{1j}(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{ij} \rceil \leq 0 \ \forall j \in \{1,2,3,4,5\}$$

4.4.4 Results

The genetic algorithm ran for 200 generations and 53,180 function evaluations. It produced a nondominated front with 74 solutions. For comparison purposes, an exhaustive search was done of all possible combinations of payload distributions across up to five modules for which each payload type was present at least once. It explored 345,861 solutions and produced a nondominated front, also with 74 solutions. Figs. 4.5 and 4.6 show that the fronts produced by the two methods are identical, meaning that the genetic algorithm successfully located all of the nondominated solutions in the design space. As expected, the lowest cost case is the monolithic case, for which the removal of a single module destroys all capability. The front contains solutions for which the removal of one or two modules does not affect the performance due to redundancies for all payloads.

The number of generations and function evaluations that were required for the GA to locate the Pareto front can be determined by examining the inverted generational distance. To calculate the inverted generational distance, the inverted nearest distance must be known. The inverted nearest distance is the Euclidean distance between a point on the reference front (the front produced by the exhaustive search) and the closest point on the front produced by the optimization, normalized by the epsilon values for each objective. The inverted generational distance is the average of the inverted nearest distance for the entire reference

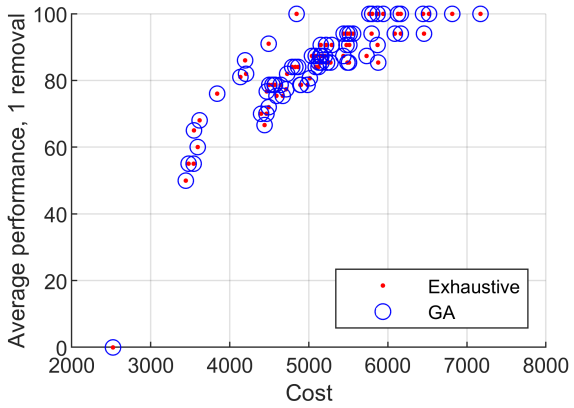


Figure 4.5: Cost versus average performance with one removal for solutions in nondominated front.

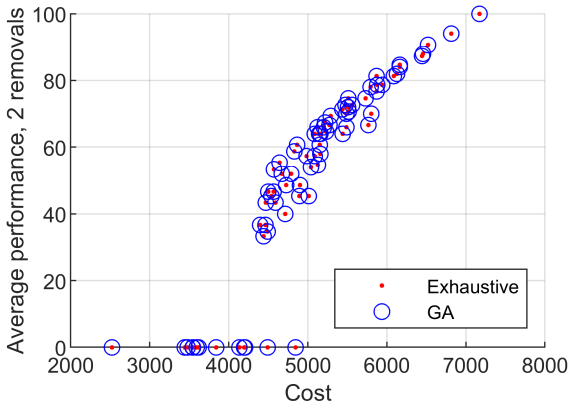


Figure 4.6: Cost versus average performance with two removals for solutions in nondominated front.

front [172]. This metric is calculated using the archive of solutions for each generation and is shown in Fig. 4.7. The inverted generational distance reaches zero (within computational tolerance) after 57 generations and 11,518 function evaluations. The Pareto front was found in 3.3% the number of function evaluations of the exhaustive search. The search continued for another 143 generations due to number of restarts until convergence specified for this optimization.

This scenario was also to compare the performance of two methods of handling constraints. The analysis as described uses the repair method to ensure that all payloads are included at least once (Repair). Additional analyses were performed in which the repairable constraint is replaced with a penalty function. Four penalty functions are considered:

1. **Replace:** The fitness values for objectives #2–4 are replaced with the number of missing payloads if not all payloads are present.

2. **Add:** The number of missing payloads is added to the fitness values for objectives #2–4 if not all payloads are present.

3. **Add 5:** The number of missing payloads times five is added to the fitness values for objectives #2–4 if not all payloads are present.

4. **Add 20:** The number of missing payloads times twenty is added to the fitness values for objectives #2–4 if not all payloads are present.

Multiple runs were performed for each constraint handling case. The inverted generational distances with respect to the front determined by full enumeration are shown in Fig. 4.8. For this payload assignment problem, the repair method outperformed the penalty function method. The repair method converged to the Pareto front and converged more quickly than any other method. Of the penalty function methods, the best performances were achieved using the 'Replace' function and the 'Add 20' function. The 'Replace' function produced solutions that matched the reference Pareto front plus five additional solutions with worse performance but lower cost than the solutions on the reference Pareto front. The 'Add 20' function produced solutions that matched the reference Pareto front. It also produced 33 solutions that were not on the reference Pareto front. These solutions had lower cost but worse performance than other solutions in the archive. Due to the high value of the penalty function when payloads are missing, none of the solutions with missing payloads dominated the reference solutions. The other two penalty function methods did not locate all of the solutions on the reference Pareto front because there were solutions with missing payloads that dominated one or more reference solutions. The 'Add 5' function missed one solution on the reference Pareto front, while the 'Add' function missed three solutions. This analysis indicates that the repair method outperforms the penalty function method for payload assignment problems. It also indicates that the choice of penalty function has a significant impact on convergence performance and on the resulting nondominated set of solutions.

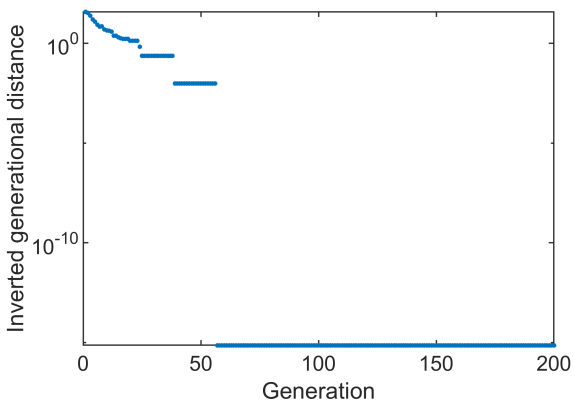


Figure 4.7: Inverted generational distance over generations.

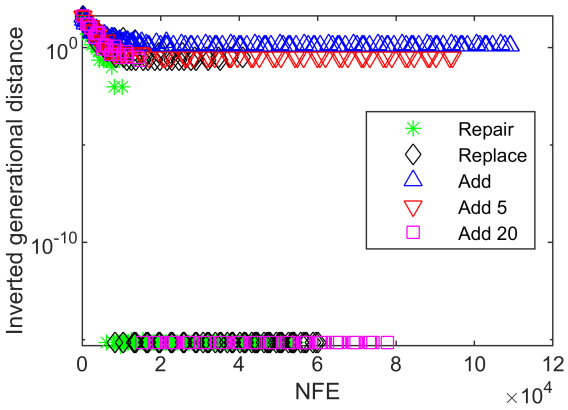


Figure 4.8: Inverted generational distance for various constraint handling methods.

4.4.5 Conclusions

This scenario demonstrated the ability of a variable-length genetic algorithm to discover Pareto-optimal candidates for payload distribution among a cluster of modules. The GA located all solutions on the Pareto front, as verified by an exhaustive search of the solution space. The GA found all candidates on the front in 3.3% the number of function evaluations of the exhaustive search. These results suggest that a variable-length GA is a good candidate for optimizing various levels of disaggregation for heterogeneous groups of satellites. An additional analysis compared the performance of constraint handling methods and determined that the repair method had the best performance for the payload assignment problem. These results will be applied to the constraint formulation and payload distribution optimization for the sea surface height monitoring scenario in Chapter 6.

Chapter 5

Applications: Resilience

5.1 Introduction

Spacecraft can fail for a variety of reasons, including launch failures, radiation, thermal stresses, and electronics failures [15]. Because the failure of even a single satellite detracts from a constellation's ability to perform its mission, it is beneficial to explore robust architectures that perform well even after the failure of one or more assets. The likelihood of failure of a satellite is not necessarily correlated with the impact the loss of that satellite has on the mission. A constellation's resilience to failure is a function of the importance of each individual satellite to the mission and of the predicted failure rates of the satellites. Conversely, mission capability can be degraded due to intentional adversarial actions. In such a case, it must be assumed that the targeted assets will be the ones whose loss most damages mission performance. Therefore, resilience to threat is a function of the importance of each individual satellite to the mission and of the ability of an adversary to damage each satellite. To capture the resilience of a system to threats, it is necessary to determine the worst-case performance of that system for some level of degradation. The same process can provide a bounds on the worst-case performance of a system with random failures by using the satellite failure rate to set the level of degradation. Previous work has measured constellation resilience by using the predicted failure rate as a surrogate for resilience [127] or by using the predicted number of satellites on orbit as a surrogate for resilience [69]. These methods do

not directly quantify the impact of satellite loss on performance. Ref. 22 suggests an equation for quantifying resilience as a function of capacity retained, recovered, and replaced but does not provide a method for calculating capacity. Cuhran, Jenkins, and Walters measured resilience by looking at probability of collision and the worst-case performance when a single satellite was lost [35]. This method works well for evaluating performance for small-scale failures but does not consider cases in which multiple satellites may be simultaneously disabled through adversarial action. Stenger performed network analysis for a degraded Iridium constellation, selecting the worst-case removals in batches of 12 by finding the satellites that appeared most often in the packet paths and removing them [141]. However, this method is not mathematically rigorous for nonuniform constellations because the problem of satellite access cannot in general be solved recursively. A method of assessing resilience that gives consideration to performance is desired. Such a method may be used to assess resilience for both satellite to ground station links and ground station to satellite links.

Disaggregation has been proposed as a method of increasing the resilience of space systems [125]. By increasing the number of satellites in a system, the impact of the failure of each satellite on mission success is lessened, allowing the constellation to survive a larger number of losses without total loss in capability. Disaggregation has the added benefit of reducing satellite size. Small satellites benefit from lower costs and opportunities for standardization not available for large satellites. However, they come with their own set of challenges. For example, the small form factor, low-cost satellites known as CubeSats are particularly susceptible to failure due to their use of COTS parts, low budgets, and high risk tolerance. A 2017 study by the Aerospace Corporation found that academic CubeSat missions failed 55% of the time, while commercial CubeSat endeavors failed 23% of the time [155]. Despite their failure rate, CubeSat constellations can enable critical space missions by providing rapid response through their rapid development times. Additionally, their small size allows them

to be launched as secondary payloads when their mission has some flexibility in the required orbital configuration. CubeSats can therefore play an important role in a disaggregated constellation if their limitations are controlled. As part of these considerations, the effect of the failure of one or more satellites on the ability of the constellation to perform its mission must be assessed. This chapter addresses the second research question: How can resilience be assessed in a tractable and comprehensive manner?

5.2 Scenario 5-a: Rideshare-based fire detection

5.2.1 Purpose and background

Rapid response is critical when dealing with natural disasters, such as hurricanes, earthquakes, and floods. The damage from these events can be mitigated by using space-based assets to provide communications or to monitor and survey damage. California wildfires are used as an example natural disaster. The economic damage from the 2018 wildfires in the State of California was estimated to be approximately $400 billion due to firefighting costs, property damage, wage loss, and health impacts [117]. A solution that improves fire detection and response time therefore has strong potential economic benefits.

Rapid response is critical when providing aid after natural disasters. Nanosatellites and rideshare opportunities can be used to reduce the time needed to deploy a satellite constellation. COTS components are available for nanosatellites, allowing for easy construction and rapid build times. Their low cost relative to traditional satellites often permits more risk tolerant behavior, reducing the amount of testing a satellite must undergo prior to deployment. This feature is especially beneficial in the event of an emergency, when there is insufficient time to build and deploy a traditional satellite constellation.

An issue in the deployment of nanosatellite constellations is getting all of the assets into orbit. In traditional spacecraft constellations, the launch costs are a relatively small portion of the overall budget—an example scenario in Ref. 165 predicts a launch cost that is 14% of the spacecraft cost. In the case of nanosatellites, however, the cost of the spacecraft itself is much smaller, generally on the order of a million dollars. Some nanosatellites are as cheap as a couple hundred thousand dollars [16]. Because launch vehicles cost tens to hundreds of millions of dollars, the use of dedicated launches for low cost missions is infeasible unless hundreds of satellites are going to the same plane [98]. However, small satellites can be launched as secondary payloads via rideshare programs for about $30,000 per kilogram [16]. Additionally, academic groups may qualify for free launch services through the Educational Launch of Nanosatellite (ELaNa) missions [120]. The downside of constellations built using rideshare alone, sometimes called ad hoc constellations, is that the irregular distribution of satellites results in large gaps in coverage compared to a symmetric constellation like a Walker constellation. Previous studies have quantified these differences, but have shown that performance can be improved through optimization of the rideshare selection. One such study randomly selected combinations of satellites from previous launches to characterize the range of performance for ad hoc constellations providing global coverage [49]. Another study used a multi-objective genetic algorithm to determine an optimal rideshare manifest for providing global coverage. A single solution from the resulting nondominated front was selected and its resiliency assessed after the optimization [61]. Finally, a method was developed for optimally manifesting secondary payloads using linear programming techniques, but this method requires that the desired orbits for the satellites be known a priori [28].

This study simulates the manifestation constellation of nanosatellites constructed of COTS parts using rideshare opportunities in a timely manner. The nanosatellites are identical and have the following subsystems:

- **Propulsion:** 2U propulsion module with a thrust of 1.25 N, a specific impulse of 235 s and a fuel mass of 1.4 kg [1]
- **Antenna:** Deployable helical antenna
- **TX/RX:** VHF downlink/UHF uplink full duplex transceiver
- **Battery:** High energy density battery array
- **EPS:** CubeSat power supply
- **Solar panels:** CubeSat Solar panels
- **ADCS:** CubeSat 3-axis reaction wheels and magnetorquers
- **Imager:** Multispectral imager with a ground sample distance of 9.6 m at 500 km [2]

With the exception of the propulsion system and the imager, the components listed above serve only to estimate the cost and mass required for the satellite and do not represent a finalized design. The propulsion system dictates the maneuvers that can be performed by the satellite, while the imager dictates the image resolution, limiting the maximum altitude of the satellites.

A set of rideshare options was simulated by taking the two line elements (TLEs) of satellites launched over a thirty day period. This sampling is meant to be an example set of launches and is not indicative of the launches that would be available for an actual mission. The results will vary based on the particular set of launches available. For comparison purposes, optimization is also performed on a Walker delta constellation with up to fifty satellites and up to twenty planes. The Walker formulation does not undergo reconfiguration. It seeks to

[1] http://www.rocket.com/files/aerojet/documents/CubeSat/MPS-130%20data%20sheet%20crop.pdf

[2] http://41.185.8.177/~cubespac/ClientDownloads/CubeADCS_3Axis_Specsheet_V1.1.pdf

minimize the total number of planes in addition to the objectives stated for the rideshare scenario. This scenario was originally presented in Ref. 116.

5.2.2 Methodology

The DISCO-Tech methodology is used to model and solve the optimization problem. The concept of resilience is introduced and a formulation for determining a system's resilience developed.

Resilience

Resilience refers to the constellation's ability to perform in the event of the loss of one or more satellites either due to system failure or due to adversarial actions. In the former case, the loss of satellites can be modelled as random and a Monte Carlo analysis used if the expected performance is desired. If the worst-case performance due to failure is needed (resilience to worst-case failure) or if the losses are due to adversarial actions, it is desirable to find the worst-case loss of coverage due to the removal of some number of assets. This process is simplified by using the access array calculated for the coverage objective described in subsection 3.6.1. Because the access array is stored after the initial coverage calculations are performed, the performance of a degraded constellation can be calculated without necessitating the recalculation of accesses. By removing one or more satellites from the access array, revisit metrics for the reduced constellation can be calculated. By performing these calculations for different combinations of removed spacecraft, it is possible to acquire bounds on the constellation performance in the event of one or more failures. It can also be determined which satellites are most critical to the mission, allowing redundant satellites to be deployed if needed. A full enumeration of all possible combinations of removals can

be performed when the number of satellites and removals is relatively small but becomes computationally infeasible for large constellations.

A straightforward method of selecting satellites for removal is to use a "greedy" algorithm: the satellite whose loss would be most damaging to performance is removed. The system is reevaluated, then the satellite whose loss would be most damaging to the new constellation is removed. The process continues until the desired number of losses has been achieved. However, this method is not mathematically rigorous for nonuniform constellations because the problem of satellite access cannot be solved recursively. To illustrate, a case in which satellites are spaced unevenly about an orbit is considered, as shown in Fig. 5.1. Assume the satellites are placed in an equatorial orbit, with ground targets on the equator. The largest gap in coverage occurs between satellites with the largest spacing. In the nominal case, the largest gap occurs between satellites C and D. If a single failure occurs, the satellite whose removal would most increase the largest gap is satellite C, denoted by an orange triangle. In this 1-removal case, the largest possible gap would occur between satellites B and D after the removal of satellite C. If the 2-removal case continues from the 1-removal case by assuming C has already been removed, the next best satellite to remove is B. Using recursion results in the largest gap being between A and D, with B and C removed. However, considering the 2-removal problem independently from the 1-removal problem shows that the best two satellites to remove are D and E, denoted by red squares. Removing these two satellites creates the largest gap in true anomaly and therefore the largest gap in coverage. In this case, the largest gap is between C and F after the removal of D and E. Because the C-F gap is larger than the A-D gap, it has been shown that relying on recursion to determine the worst-case removals yields an incorrect answer.

It is therefore beneficial to develop a methodology for determining the satellites whose removal is most damaging to the constellation performance. If the number of satellites is small

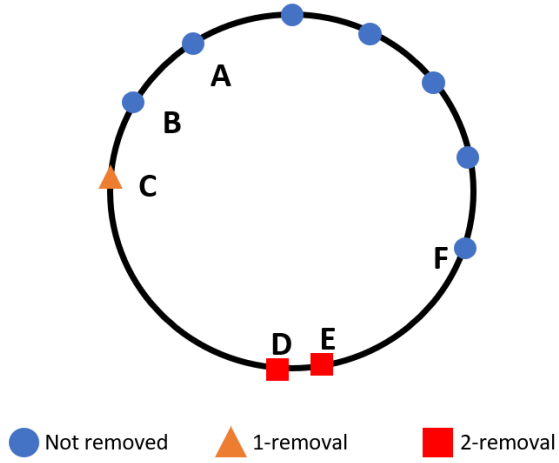


Figure 5.1: Representative nonuniform satellite distribution.

and the number of predicted removals/failures is small, it is simple to check every combination of removals to determine the most damaging case. However, because the problem grows combinatorially, evaluating all combinations directly for large problems quickly becomes infeasible. For example, Stenger considered 12-removal, 24-removal, and 36-removal cases for the 66 satellite Iridium constellation [141]. The smallest of these cases has 4.92×10^{12} combinations that would need to be evaluated.

Mathematical programming is used to solve the problem of finding the worst combination of asset removals, which can be formulated as a mixed integer linear program (MILP). The exact formulation will depend on the metric used in the optimization. This subsection develops a formulation for finding the combination of removals that maximizes the longest period of time any target is without coverage (maximum revisit time). Additional objective formulations are described in section 5.3 and in Appendix B. The targets are assumed to be ground stations, and the assets are assumed to be satellites.

All of the formulations require the decision variable $x_j \in \{0,1\} \forall j = 1, 2, \ldots, n_s$, a binary satellite inclusion variable that is one if the j-th satellite is active and zero otherwise. The

degraded case assumes that n_r satellites are lost through system failure or through adversarial action. To find the correct number of removals, it is required that

$$\sum_{j=1}^{n_s} x_j = n_s - n_r \tag{5.1}$$

It is then necessary to determine how many satellites are available to each station at any given time. The access sum, $\overline{A}_{ik}$, gives the number of satellites available to the k-th ground station at its i-th time step,

$$\overline{A}_{ik} = \sum_{j=1}^{n_s} A_{ijk} x_j \tag{5.2}$$

where A_{ijk} is an element of the access array discussed in the coverage analysis subsection, subsection 3.6.1.

It is then necessary to calculate whether a sufficient number of assets are available for access at each period. A successful access requires that at least n_c assets be in view of a station simultaneously. The calculation of this access requires the introduction of a new binary variable, $Y_{ik} \in \{0, 1\}$. Y_{ik} is one if the required number of assets are accessible by station k at time step i and zero otherwise. To force Y_{ik} to take the appropriate value, two constraints are introduced. To ensure that Eq. (5.4) is satisfied whenever $Y_i k$ is one, $M = n_s - n_c + 1$.

$$Y_{ik} \leq \frac{\overline{A}_{ik}}{n_c} \qquad \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \tag{5.3}$$

$$M Y_{ik} \geq \overline{A}_{ik} - n_c + 1 \qquad \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \tag{5.4}$$

Eq. (5.3) ensures that Y_{ik} is zero if insufficient satellites are available. Eq. (5.4) ensures that Y_{ik} is one if at least n_c satellites are available.

The remainder of the problem formulation is dependent upon the metric used for the optimization and whether the problem is a minimization or maximization problem.

The maximization of the maximum revisit time across all ground stations requires the definition of an accumulator variable $a_{ik} \in \mathbb{R}_{\geq 0}$. The accumulator variable should count the amount of time at each step since the end of the previous pass. During a pass and immediately after the pass ends, the accumulator should be zero. M_2 in Eqs. (5.5)–(5.8) must be large enough that the constraints are satisfied whenever Y_{ik} is one. The most conservative value for M_2 is the length of the scenario plus a small constant. However, using smaller values to aid convergence is encouraged if it is guaranteed that no gap will ever exceed the value chosen for M_2.

The constraints needed to force a_{ik} to take the appropriate value are slightly different for the first time step than for the rest of the scenario. a_{1k} has the constraints

$$a_{1k} \geq \Delta T_{1k} - M_2 Y_{1k} \; \forall \; k = 1, 2, \ldots, n_g \tag{5.5}$$

$$a_{1k} \leq \Delta T_{1k} + M_2 Y_{1k} \; \forall \; k = 1, 2, \ldots, n_g \tag{5.6}$$

These constraints ensure that a_{1k} will be equal to the length of the first time step if there is no access when the scenario begins. The constraints for the rest of the time period are

$$a_{ik} \geq a_{(i-1)k} + \Delta T_{ik} - M_2 Y_{ik} \; \forall \; i = 2, 3, \ldots, n_t - 1, \;\; k = 1, 2, \ldots, n_g \tag{5.7}$$

$$a_{ik} \leq a_{(i-1)k} + \Delta T_{ik} + M_2 Y_{ik} \; \forall \; i = 2, 3, \ldots, n_t - 1, \;\; k = 1, 2, \ldots, n_g \tag{5.8}$$

Likewise, these constraints ensure that a_{ik} will be equal to the previous accumulator value plus the time step if there is no access at the current time. Finally, a_{ik} must be zero if there is access at the current time, so

$$a_{ik} \leq M_2(1 - Y_{ik}) \; \forall \; i = 1, 2, \ldots, n_t - 1, \;\; k = 1, 2, \ldots, n_g \tag{5.9}$$

The length of the largest gap is equal to the largest value of a_{ik}. To find this value, a variable $a_{\max} \in \mathbb{R}_{\geq 0}$ is introduced. The remainder of the problem formulation varies based on whether the goal is to maximize the largest gap (maximizing the maximum revisit time) or minimize the largest gap (minimizing the maximum revisit time). The latter case is discussed in Appendix B. The maximization problem requires the inclusion of additional binary variables δ_{ik}. Because the goal is to maximize $a_{\max}$, there must be an upper bound on $a_{\max}$ to prevent it growing unbounded. Therefore, it is required that $a_{\max}$ is less than or equal to exactly one of the values of a. This formulation will drive $a_{\max}$ to the largest value of a and can be enforced with the constraints

$$a_{\max} \leq a_{ik} + (1 - \delta_{ik})M_2 \;\forall\; i = 1, 2, \ldots, n_t - 1,\; k = 1, 2, \ldots, n_g \tag{5.10}$$

$$\sum_{i=1}^{n_t-1} \sum_{k=1}^{n_g} \delta_{ik} = 1 \tag{5.11}$$

The full formulation for the maximization problem is

Minimize:

$$-a_{\max}$$

such that:

$$x_j \in \{0, 1\} \qquad \forall\; j = 1, 2, \ldots, n_s$$

$$Y_{ik} \in \{0, 1\} \qquad \forall\; i = 1, 2, \ldots, n_t,\; k = 1, 2, \ldots, n_g$$

$$a_{ik} \in \mathbb{R}_{\geq 0} \qquad \forall\; i = 1, 2, \ldots, n_t - 1,\; k = 1, 2, \ldots, n_g$$

$$a_{\max} \in \mathbb{R}_{\geq 0}$$

$$\delta_{ik} \in \{0, 1\} \qquad \forall\; i = 1, 2, \ldots, n_t - 1,\; k = 1, 2, \ldots, n_g$$

$$\sum_{j=1}^{n_s} x_j = n_s - n_r$$

$$
\begin{aligned}
Y_{ik} &\leq \frac{\sum_{j=1}^{n_s} A_{ijk} x_j}{n_c} && \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g \\
MY_{ik} &\geq \sum_{j=1}^{n_s} A_{ijk} x_j - n_c + 1 && \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g \\
a_{1k} &\geq \Delta T_{1k} - M_2 Y_{1k} && \forall\ k = 1, 2, \ldots, n_g \\
a_{1k} &\leq \Delta T_{1k} + M_2 Y_{1k} && \forall\ k = 1, 2, \ldots, n_g \\
a_{ik} &\geq a_{(i-1)k} + \Delta T_{ik} - M_2 Y_{ik} && \forall\ i = 2, 3, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{ik} &\leq a_{(i-1)k} + \Delta T_{ik} + M_2 Y_{ik} && \forall\ i = 2, 3, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{ik} &\leq M_2 (1 - Y_{ik}) && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{\max} &\leq a_{ik} + (1 - \delta_{ik}) M_2 && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
\sum_{i=1}^{n_t - 1} \sum_{k=1}^{n_g} \delta_{ik} &= 1
\end{aligned}
$$

This formulation has a total of $(2n_t - 1) \times n_g + n_s$ binary variables and $(n_t - 1) \times n_g + 1$ continuous variables.

Problem size

State-of-the-art MILP solvers can handle hundreds of thousands of variables, but the ability of a solver to find a solution is dependent on more factors than simply the number of variables [87]. The distance between the relaxed and actual solutions, the number of constraints, and the ability of the solver to quickly find a feasible solution are only some of the factors that affect solve time.

There are a couple of strategies that can be employed to reduce computation time. The first is to tighten the values of M and M_2 as much as possible. Using separate values for each ground station, rather than a single value across all stations, can allow for a greater reduction. M can be set to the largest number of satellites that each station sees simultaneous in the nominal case. M_2 can be reduced if an upper bound on the gap size for each station can

be predicted. Finally, some decision variables can be set *a priori*. For example, any interval where the nominal case did not have access to a sufficient number of assets will preclude the reduced case from having a sufficient number of assets, allowing the corresponding values of Y_{ik} to be set to zero. Similarly, if the nominal case had sufficient assets such that removing the maximum number of assets would not affect access during that period, the corresponding values of Y_{ik} can be set to one.

Another method for decreasing runtime is to consider each ground station separately. Because the runtime of an MILP solver will increase in a non-polynomial fashion with number of variables, it is generally faster to run several small programs than one large one. Meaning, one program with 100,000 variables will generally be slower than one hundred programs with one thousand variables each. The results can then be combined by either choosing the satellites whose removals occur most often or by evaluating each set of removals against the objective function and choosing the set with the worst performance. The latter method will, in the case of the maximization or minimization over the ground stations (as with maximum revisit time), give the same answer as the general formulation and is therefore used.

The number of ground stations and time steps both have a more significant direct impact than the number of satellites, though the number of time steps is a function of the number of satellites. It is possible to reduce the size of the problem by limiting the number of ground stations or the total duration of the scenario. The problem size is assessed for two examples.

- Iridium constellation ground coverage: The Iridium constellation uses 66 satellites to provide global coverage. An analysis with 100 points scattered around the globe is assumed. It is also assumed that each station experiences two accesses per satellite per day on average. These approximations give $n_t = 264$, $n_g = 100$, and $n_s = 66$. The problem would have 52,766 binary variables and 26,301 continuous variables for

a single day. However, if the number of ground points can be reduced to 10 and still provide a representative set of ground points, the problem would have only 5,336 binary variables and 2,631 continuous variables per day.

- Iridium constellation crosslink: The satellites can be treated simultaneously as the "ground stations" to evaluate the ability of the Iridium constellation to maintain crosslinking under degradation. Since the constellation is uniform, only a single orbital period need be considered. Because the satellites only crosslink with nearby satellites, the number of rise-sets for each satellite will be small – approximately 4. These approximations give $n_t = 8$, $n_g = 66$, and $n_s = 66$. The resulting problem has only 1,056 binary variables and 463 continuous variables.

Constellations with repeating ground tracks or whose formulations permit short analysis times will benefit most from this formulation.

5.2.3 Optimization formulation

Mission definition

The optimization selects a set of launches, assigns satellites to each selected launch, and sets a reconfiguration for each launch by setting the changes in orbital elements. The satellites' orbital elements will be the orbital elements of the launch plus the change in orbital elements. The orbital elements for each Earth-orbiting rideshare launch are shown in Table 5.1.

Table 5.1: Rideshare orbital elements

Launch	a(km)	e	$i(^\circ)$	$\omega(^\circ)$	$\Omega(^\circ)$	$\nu(^\circ)$
1	6,823	0.0018	92.9	118.3	253.5	159.8
2	6,823	0.0018	92.9	118.6	253.5	159.9
3	6,823	0.0017	92.9	124.2	253.5	156.2
4	6,823	0.0018	92.9	124.6	253.5	156.3

5	6,966	0.0014	97.9	153.2	158.3	10.8
6	6,966	0.0014	97.9	151.1	158.3	8.8
7	28,241	0.0117	55.0	176.3	156.9	125.5
8	28,243	0.0116	55.0	176.4	156.9	124.8
9	7,090	0.0091	98.6	337.4	340.2	241.5
10	15,700	0.5808	55.0	172.5	153.7	221.1
11	15,531	0.5723	26.9	195.9	240.4	154.9
12	7,161	0.0011	98.6	165.3	339.7	177.0
13	6,673	0.0021	51.6	340.3	237.9	19.7
14	6,837	0.0019	91.9	59.6	251.1	163.6
15	6,975	0.0032	97.7	175.0	158.0	216.6
16	6,784	0.0008	51.6	35.9	237.1	315.8
17	42,133	0.0010	0.0	96.4	95.9	188.3
18	28,820	0.0089	54.9	4.6	156.8	179.5

The satellites will be evenly distributed in true anomaly around the orbit. Up to twenty launch opportunities may be selected. It is possible that multiple planes of satellites may be deployed from a single launch by assigning multiple orbital element changes to one launch. The decision variables for the optimization are shown in Table 5.2.

Table 5.2: Fire detection decision variables.

DV	Description	Lower limit	Upper limit	Type
x_{i1}	Orbit selection index	1	18	$\mathbb{Z}$
x_{i2}	Satellites in the plane	1	10	$\mathbb{Z}$
y_{i1}	Change in semimajor axis (Δa), m	-1×10^6	1×10^6	$\mathbb{R}$
y_{i2}	Change in eccentricity (Δe)	-0.1	0.1	$\mathbb{R}$
y_{i3}	Change in inclination (Δi), °	-5	5	$\mathbb{R}$
y_{i4}	Change in argument of periapsis ($\Delta\omega$), °	-10	10	$\mathbb{R}$
y_{i5}	Change in RAAN ($\Delta\Omega$), °	-5	5	$\mathbb{R}$
y_{i6}	Change in true anomaly ($\Delta\nu$), °	-180	180	$\mathbb{R}$
$i_{\max}$	Number of planes	1	20	$\mathbb{Z}$
for i	plane ID	1	$i_{\max}$	$\mathbb{Z}$

Optimizer

The following optimizer parameters are used.

- **Optimizer:** Variable-length genetic algorithm with epsilon boxing

- **Initial population:** 200
- **Tournament competitors:** 4
- **Mutation rate:** 5%
- **Length change rate:** 5%
- **Expansion rate:** 0%
- **Nondominated generations for convergence:** 10
- **Number of restarts:** 10
- **Maximum archive size:** 300
- **Epsilon boxes:** [0.01, 0.01, 2, 0.01]
- **Objective limits:** [Inf, Inf, Inf, Inf]

Dynamics

The following dynamics parameters are used.

- **Propagator:** Fast J4 propagator
- **Central body:** Earth
- **Scenario epoch:** Oct 9, 2018, 00:00
- **Scenario duration:** 30 days
- **Propagation time step:** 86.4 sec
- **Satellite pointing:** Aimed

Constraints

Three constraints are used in this scenario.

Constraint #1, number of satellites: No more than fifty satellites can be used in the

constellation. Adapted from Eq. (3.37) in subsection 3.5.3, the constraint equation is

$$g_1(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i2} - 50 \leq 0 \tag{5.12}$$

This constraint is not treated as repairable. A penalty value of $9,999(50 - \sum_{i=1}^{i_{\max}} x_{i2})$ is assigned for solutions violating the constraint.

Constraint #2, reachability: The satellites must have sufficient fuel to make the transition from their initial orbits to their final orbits. The final orbit is reachable if the Δv available on a satellite is greater than or equal to the amount needed to make the transition, which can be approximated using the process described in subsection 3.5.5. The spacecraft estimated dry mass is 4.35 kg based on the COTS component mass. Since the propulsion module has a specific impulse of 235 s and a maximum fuel mass of 1.4 kg, the maximum allowable Δv is 643 m/s. Adapted from Eq. (3.41), the reachability constraint requires that

$$g_{2i}(\mathbf{x}_i, \mathbf{y}_i) = \Delta v_i(\mathbf{x}_i, \mathbf{y}_i) - 643 \text{ m/s} \leq 0 \ \forall i = 1, 2, \ldots, i_{\max} \tag{5.13}$$

This constraint is repairable. If the constraint is violated, the orbital changes are reduced in an attempt to satisfy the constraint. As a first effort, the ratio between the allowable and needed Δv is found and used as a multiplier to scale the changes in RAAN and inclination. These parameters are targeted first due to their large fuel cost. If that fails, all orbital changes are gradually reduced by 10% until the constraint is satisfied. The process is repeated for each satellite.

Constraint #3, eccentricity: The final eccentricity of each orbit must be greater than or equal to zero. If $e_{x_{i1}}$ is the eccentricity of the launch vehicle to which the i-th set of satellites

is assigned, the constraint is

$$g_{3i}(\mathbf{x}_i, \mathbf{y}_i) = -e_{x_{i1}} - y_{i2} \leq 0 \ \forall i = 1, 2, \ldots, i_{\max} \tag{5.14}$$

This constraint is repairable. If constraint is violated, y_{i2} is modified such that the final eccentricity is equal to 0.001.

Objectives

This scenario has four objectives.

Objective #1, minimize average time average gap: Coverage is calculated over points evenly spaced every 100 miles in the State of California. Access requires a single satellite at an elevation angle of at least 5°. The ground sample distance (GSD) of the image must be less than or equal to 25 m, which imposes a range limitation on access. The calculation of time average gap is as described in subsection 3.6.1. The time average gap (TAG) is calculated for each ground point, then averaged. This objective quantifies typical access performance for the scenario. The epsilon box for this objective is set at 0.01 days, or 864 seconds. If there are n ground points, g_{km} is the length of the m-th gap time for the k-th ground point, n_{gk} is the number of gaps for the k-th ground point, and D is the scenario duration, the objective function can be found by averaging Eq. (3.64) across all targets:

$$f_1(\mathbf{x}, \mathbf{y}) = \frac{\sum_{k=1}^{n} \sum_{m=1}^{n_{gk}} g_{km}(\mathbf{x}, \mathbf{y})^2}{nD} \tag{5.15}$$

Objective #2, minimize maximum revisit time: The access calculation is the same as for TAG. The calculation of maximum revisit time is as described in subsection 3.6.1. Maximum revisit time is calculated for each ground point, then the maximum value found

to get the largest gap for any ground point at any time in the scenario. This objective quantifies the worst access performance for the scenario. The epsilon box for this objective is set at 0.01 days, or 864 seconds. The objective function can be found by maximizing Eq. (3.62) across all targets:

$$f_2(\mathbf{x},\ \mathbf{y}) = \max_k \left(\max_m g_{km}(\mathbf{x},\ \mathbf{y}) \right) \tag{5.16}$$

Objective #3, minimize number of satellites: This objective is set to explore the design space and examine the performance of constellations of various sizes. The epsilon box for this objective is set at two satellites. The objective function is

$$f_3(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i2} \tag{5.17}$$

Objective #4, minimize degraded maximum revisit time: The degraded constellation performance is evaluated assuming a loss of 20% of the satellites. Degraded performance is evaluated over a period of ten days. An MILP solver is used to determine the combination of satellites that gives the worst performance with respect to maximum revisit time. This process is described in subsection 5.2.2. The epsilon box for this objective is set at 0.01 days, or 864 seconds. If g'_{km} is the length of the m-th degraded gap time for the k-th ground point, the objective function can be found by maximizing Eq. (3.62) across all targets:

$$f_4(\mathbf{x},\ \mathbf{y}) = \max_k \left(\max_m g'_{km}(\mathbf{x},\ \mathbf{y}) \right) \tag{5.18}$$

Mathematical summary

The mathematical formulation for the optimization is shown below.

Minimize

$$f_1(\mathbf{x},\,\mathbf{y}) = \frac{\sum_{k=1}^{n}\sum_{m=1}^{n_{gk}} g_{km}(\mathbf{x},\,\mathbf{y})^2}{nD}$$

$$f_2(\mathbf{x},\,\mathbf{y}) = \max_k \Big(\max_m g_{km}(\mathbf{x},\,\mathbf{y})\Big)$$

$$f_3(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i2}$$

$$f_4(\mathbf{x},\,\mathbf{y}) = \max_k \Big(\max_m g'_{km}(\mathbf{x},\,\mathbf{y})\Big)$$

such that

$$i_{\max} \in [1, 20] \cap \mathbb{Z}$$

$$x_{i1} \in [1, 18] \cap \mathbb{Z}\ \forall i = 1, 2, \ldots, i_{\max}$$

$$x_{i2} \in [1, 10] \cap \mathbb{Z}\ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i1} \in [-10^6, 10^6]\ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i2} \in [-0.1, 0.1]\ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i3} \in [-5, 5]\ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i4} \in [-10, 10]\ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i5} \in [-5, 5]\ \forall i = 1, 2, \ldots, i_{\max}$$

$$y_{i6} \in [-180, 180]\ \forall i = 1, 2, \ldots, i_{\max}$$

$$g_1(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i2} - 50 \leq 0$$

$$g_{2i}(\mathbf{x}_i, \mathbf{y}_i) = \Delta v_i(\mathbf{x}_i, \mathbf{y}_i) - 643\ \text{m/s} \leq 0\ \forall i = 1, 2, \ldots, i_{\max}$$

$$g_{3i}(\mathbf{x}_i, \mathbf{y}_i) = -e_{x_{i1}} - y_{i2} \leq 0\ \forall i = 1, 2, \ldots, i_{\max}$$

5.2.4 Results

The rideshare simulation produced a nondominated front with 31 results. The front is shown in Fig. 5.2, where the points are nondominated across all four objectives. Fig. 5.3 shows only the points that form the nondominated front in two objectives – the number of satellites and one of the revisit metrics. The TAG of the nondominated solutions ranges from 1 hour for the larger constellations to 11.2 hours for a single satellite. The maximum revisit time over all points takes values between 4.4 and 12.4 hours. The degraded maximum revisit time over all points takes values between 5.2 and 11.8 hours, discounting the single-satellite case. Diminishing returns are experienced as satellites are added beyond the tenth. Indeed, the objective values change very slightly between 20 and 35 satellites.

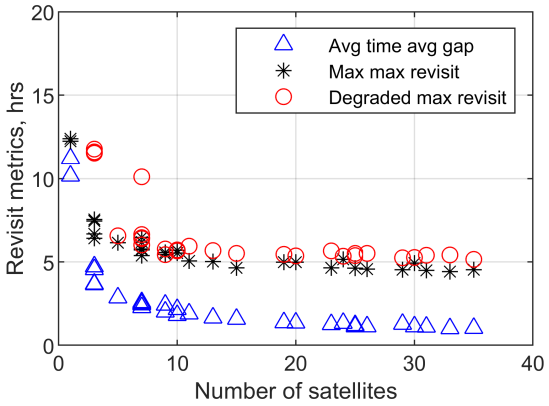


Figure 5.2: Nondominated front for rideshare-launched constellation.

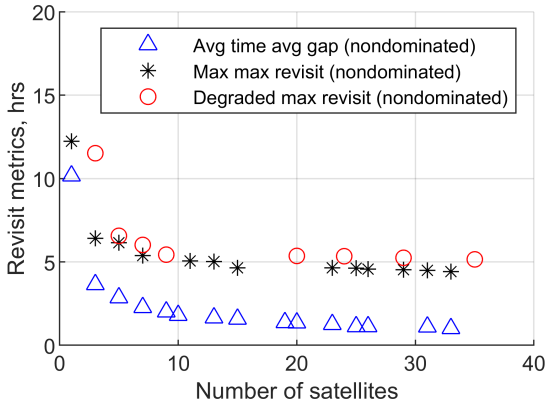


Figure 5.3: Nondominated front in two objectives for rideshare-launched constellation.

Fig. 5.4 shows the number of launches used by each solution. The points that form the nondominated front in number of satellites and average TAG are marked with a circle, while points that are dominated in number of satellites and average TAG (but nondominated across all four objectives) are marked with an x. The theoretical FireSat-II example in Ref. [165] requires a revisit time of eight hours to identify nascent forest fires. The imaging capability provided by the rideshare constellations is sufficient to meet these requirements.

The rideshare constellation performance is inferior to proposed constellation designs such as the FUEGO program, which achieves 25 minute revisit times using dedicated launches [50].

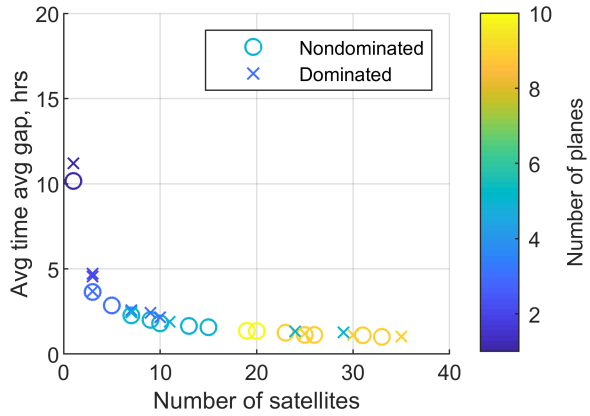


Figure 5.4: Average time average gap for rideshare-launched constellation.

There are solutions for which the nominal maximum revisit time, which is calculated over a 30 day period, is greater than the degraded maximum revisit time, which is calculated over a 10 day period. This discrepancy occurs when the largest gap in coverage does not occur within the first ten days of the simulation. To mitigate this occurrence, the analysis period for the resilience calculation can be centered on the time of the nominal maximum gap in coverage. Alternately, the resilience calculation period can be expanded at the expense of additional computation time.

The stagnation of the values with increasing numbers of satellites highlights the critical flaw in rideshare constellations. Because the initial launches are fixed in such a way that may not be beneficial to the rideshare mission, the resulting constellation can have large gaps in coverage when the rideshare orbits do not overlap in a fortuitous manner. The ability to maneuver the satellites helps to mitigate the problem, but the high fuel cost to enact a change in orbital plane impedes the constellation's ability to achieve the uniform formation often used in constellation design.

There are two ways to increase the performance of the rideshare constellation. The first is to have a greater number of rideshare opportunities available. By permitting the satellites to be launched over a longer time period, more rideshares become available, increasing available orbit diversity. However, spreading the launch of the constellation over a longer period of time decreases the overall life of the constellation, since the time from when the constellation is fully populated to when the first satellite reaches the end of its life is decreased. Another method for increasing constellation performance and spacing between orbits is to increase the maneuvering capability of the satellites. Maneuvering can be improved by either increasing the amount of fuel on board the satellites or by using a low-thrust, high I_{sp} electric propulsion system. The latter case increases the overall maneuverability, but requires more time for the constellation to reach its final configuration. Other techniques for orbit spacing, such as differential drag or temporary orbit raising to vary the effects of Earth's oblateness on the satellites, can also be used.

Downselection

The candidate solutions provide trade-offs among the objectives, but a single solution must ultimately be selected. A common approach in choosing a solution is to provide the candidate solutions to an expert and have them choose based off of their experience. Another option is to identify all solutions that meet the requirements and select the solution that minimizes a single objective, such as cost. A final option is to employ multi-criteria decision making (MCDM) strategies that mathematically assess the trade-offs between objectives and attempt to locate a "Pareto-compromise design", a solution providing trade-offs that are mutually agreeable among all objectives [64]. This analysis discards solutions with fewer than three satellites, then uses MCDM to choose from the remaining solutions.

The selection process described in Ref. 64 is applied to the remaining solutions. The process

scales each objective from zero to one. Objectives are treated one at a time. The curve formed by the current objective and an aggregation of the remaining objectives is mapped to a quarter circle while maintaining ordinal positions. The point on the quarter circle with a slope of -1 is chosen as the Pareto-compromise point and mapped back into the original objective space. This value determines the Pareto-compromise objective value for the current objective. The process is repeated for all objectives until the Pareto-compromise objective vector is known. This objective vector may not match any known solutions, so the existing solution with the lowest mean-square-error to the Pareto-compromise objective vector is chosen. The reshaping process is shown for the normalized number of satellites in Fig. 5.5. The normalized objective value for the Pareto-compromise point is marked in black.

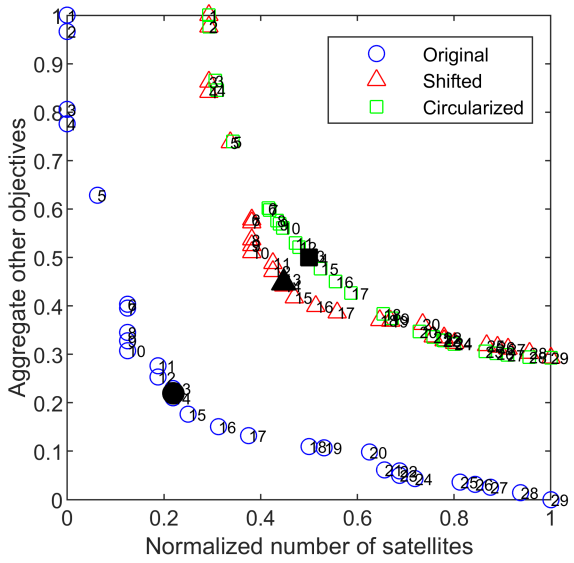


Figure 5.5: MCDM Pareto-compromise determination process for number of satellites.

The solution that most nearly approximates the Pareto-compromise objective vector is a solution of ten satellites, shown in Fig. 5.6. The constellation consists of two sets of near-polar orbits spaced about ninety degrees apart in RAAN, plus an orbit near 50 degrees in inclination. Each orbit has two satellites. During the thirty day simulation time, the polar

orbits maintain similar relative positions, but the relative positions of the 50 degree orbit varies with respect to the polar orbits. The selected solution has an average TAG of 1.8 hours and a maximum revisit time of 5 hours. The maximum revisit time increases by an hour for the degraded constellation.

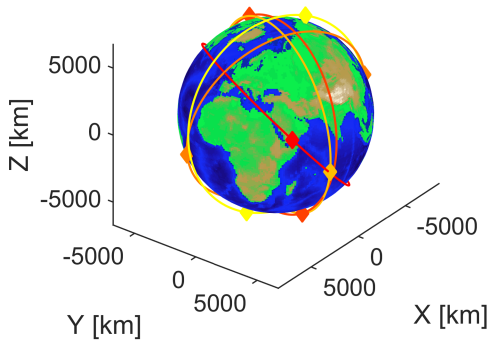


Figure 5.6: Selected constellation of 10 satellites.

Walker comparison

The performances of the rideshare constellations are compared to that of the Walker constellations optimized using the genetic algorithm. The nondominated front for the Walker case is shown in Fig. 5.7, where the points are nondominated across all five objectives. Fig. 5.8 shows only the points that form the Pareto front in two objectives – the number of satellites and one of the revisit metrics. As expected, the Walker constellations offer superior performance over the rideshare constellations for the same number of satellites. A Walker constellation of four satellites has comparable performance to a rideshare constellation of 15–20 satellites. Furthermore, satellites can be added to the Walker constellation to improve coverage until continuous coverage is reached, whereas the rideshare constellation has unfillable gaps due to the relative placement of the rideshare orbits. However, the cost of launching four satellites on dedicated rides is significantly greater than the cost of the addi-

tional satellites needed for the rideshare constellation. The cost for the components listed in the previous section is $283K for everything except the propulsion system, which is still in development and does not have a published price. If the total cost is approximately $400K with the propulsion system, the satellite cost would be $6M for the rideshare constellation and $1.6M for the Walker constellation. Neither price includes the cost of testing, integration, software development. The Walker constellation would require two to four launches to LEO, a cost of $36.8M-73.6M using Pegasus XL rockets [165]. Conversely, with a $30K per kilogram rideshare launch cost and a spacecraft weighing about 10 kg, the rideshare launch cost is only $4.5M. Therefore, if the performance limitations of the rideshare constellation are acceptable, a constellation can be developed for about a quarter of the cost of a traditional Walker constellation. However, the cost benefit of rideshare launches will decrease as launch costs decrease, either through the development of low-cost launch vehicles or through the reuse of launch vehicles. Fires in the state of California cause billions of dollars in damage each year, so the low cost of a fire detection constellation has the potential to pay for itself many times over.

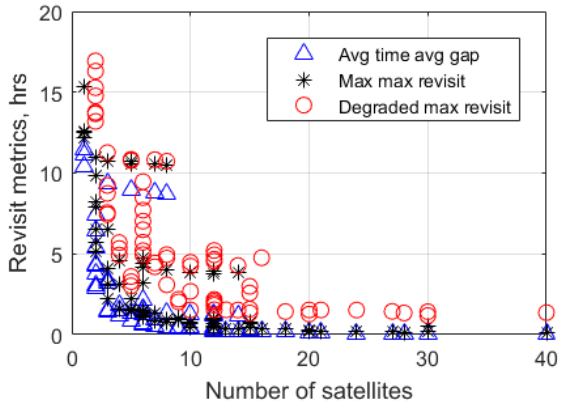


Figure 5.7: Nondominated front for Walker constellation.

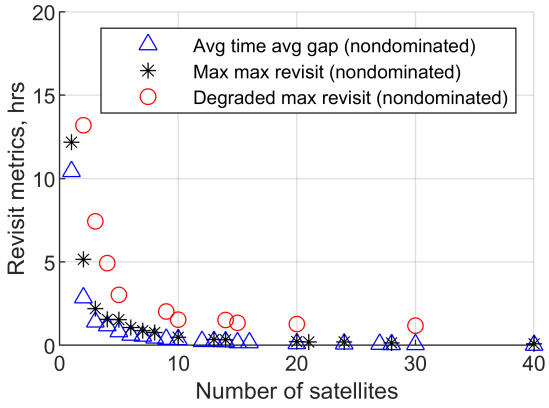


Figure 5.8: Nondominated front in two objectives for Walker constellation.

The resilience was assessed with a loss of 20% of the satellites to simulate the high failure

rate of CubeSats. The gap between the nominal maximum revisit time and the degraded maximum revisit time is about 1.5 hours for rideshare solutions with 10 or more satellites. If the number of lost satellites is independent of the total number of satellites, the gap between the nominal maximum revisit time and the degraded maximum revisit time decreases as the number of satellites increases. A fixed number of losses may be expected to occur when the losses are due to intentional adversarial action rather than system failure. Likewise, the gap between the nominal maximum revisit time and the degraded maximum revisit time decreases when the average value for degraded revisit time is found using a Monte Carlo analysis, quantifying the resilience to expected failure. The difference between the nominal and degraded maximum revisit times over a ten day period is shown in Fig. 5.9 for the case in which 20% of the satellites have been lost, the case in which three satellites have been lost, and a Monte Carlo analysis for which 20% of the satellites are removed randomly. One thousand runs were performed in the Monte Carlo analysis for each member of the archive. Although little improvement is seen in the nominal performance and the worst-case 20% loss performance beyond 15 satellites, the degraded performance improves with increasing numbers of satellites for fixed numbers of satellite losses. This result implies that increased numbers of satellites do not improve a constellation's resilience to worst-case failure (since the ratio of failed satellites remains the same regardless of the total number of satellites) but do improve the constellation's resilience to threat, which assumes a fixed number of losses. Also, the average difference between nominal performance and degraded performance decreases when the removals due to failure are chosen randomly, indicating that the resilience to expected failure increases with increasing numbers of satellites.

Cost (represented by the number of satellites) increases with increasing performance for both the rideshare and the Walker constellation cases. This scenario did not vary the mass or complexity of the satellites. Degraded performance also increases with increasing cost for

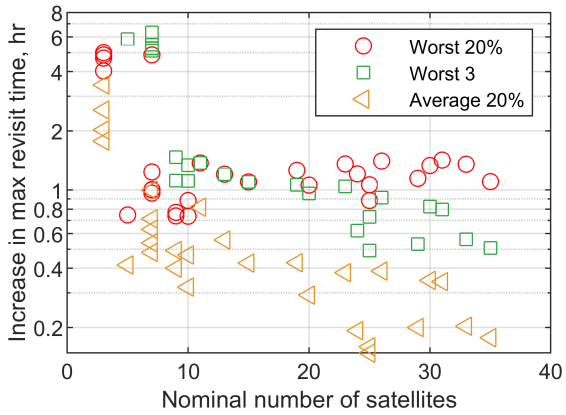


Figure 5.9: Difference between nominal and degraded maximum revisit time over a ten day period.

small numbers of satellites, though the increase in degraded performance generally corresponds to an increase in nominal performance. Resilience to threat increases with increasing costs. There are trade-offs between nominal performance and degraded performance for fixed numbers of satellites when the number of satellites in the nominal constellation is low, as indicated by the presence of multiple solutions with a range of degraded revisit times. It is therefore possible to sacrifice nominal performance for improved performance in the event of system loss. This result is particularly important when multiple solutions meet a given threshold for nominal performance. In that case, it may be advantageous to choose a solution that has better degraded performance rather than one that has better nominal performance.

5.2.5 Conclusions

This analysis included the derivation of a novel methodology for quantifying the performance of a satellite constellation in the event of the loss of one or more spacecraft. This resilience methodology was applied to the problem of designing a resilient, rideshare-based constellation for the purpose of performing fire detection over the State of California. Thirty-one nondominated rideshare solutions were found. The minimum time average gap of the solu-

tions was one hour, which provides sufficient coverage to improve fire detection capabilities. The performance of the rideshare solutions was compared to the performance of a set of Walker constellations. Although the Walker constellations produced superior coverage for a set number of satellites, the low cost of CubeSats compared to the cost of launch vehicles allow for the creation of a rideshare constellation at a quarter of the cost of a Walker constellation with comparable performance.

5.3 Scenario 5-b: Resilient ground station expansion

5.3.1 Purpose and background

The Global Positioning System (GPS) provides military and civilian users with real-time positioning information. GPS consists of a space segment of 24 satellites (plus some spares) and a ground segment of control stations, monitor stations, and ground antennas. The master control station in Colorado Springs, CO consolidates data from the monitoring stations and controls satellite operations. The monitor stations have high-precision atomic clocks and receivers that continuously measure the distance to the GPS satellites. This information is processed and used to calculate perturbations in the satellites' orbits. The ground antennas send command information and orbital knowledge corrections based on the monitor station measurements to the satellites under the direction of the master control station. The four ground antennas are located at Ascension Island, Diego Garcia, Kwajalein, and Cape Canaveral [75].

There were originally five GPS monitor stations. As part of the modernization effort, a station was added at Cape Canaveral, FL in 2001. Six more stations were added in 2005, and an additional five were added in 2006 for a total of 17 monitor stations. The locations

are shown in Table 5.3 [75].

Table 5.3: GPS monitor stations.

Original	**2001**	**2005**	**2006**
Hawaii	Cape Canaveral, FL	Adelaide, Australia	Fairbanks, AK
Colorado Springs, CO		Buenos Aires, Argentina	Osan, South Korea
Ascension Island		Manama, Bahrain	Papeete, Tahiti
Diego Garcia		Quito, Ecuador	Pretoria, South Africa
Kwajalein		Washington, D.C.	Wellington, New Zealand
		Hermitage, United Kingdom	

GPS has become increasingly important for both governmental and civilian activities. According to a 2011 study by NDP Consulting, 3.3 million jobs in the US rely on GPS. This study estimates that the economic impact of GPS disruption would be a loss of approximately $96 billion per year. The industries most affected by such a loss would be agriculture, transportation, and engineering construction [128]. Military systems, including satellites, ground vehicles, and aircraft, are reliant upon GPS. A loss of GPS capability would therefore harm both civilian activities and military efforts. As a result, it is crucial that the GPS architecture be resilient to threats. This resilience has several aspects, including resilience to spoofing, the ability to operate when satellites are lost, and the ability to operate when ground stations are lost. This study focuses on the latter aspect of resilience by examining the resilience of the monitor station network and optimizing the placement of additional monitor stations to improve resilience.

5.3.2 Methodology

The DISCO-Tech methodology is used to model and solve the optimization problem. The resilience framework developed in Section 5.2 is expanded to use the average time average gap (TAG) as the objective. This analysis does not consider geopolitical boundaries in the placement of the ground stations but could be extended to do so. Because the GPS satellites

are in repeater orbits with a period of about twelve hours, only a single day must be simulated to characterize the performance.

Resilience

For this scenario, the average time average gap for a degraded solution is used as the objective. The initial formulation for this objective is the same as the maximum revisit time objective described in Section 5.2 up to Eq. (5.4). The remaining equations must be developed for this specific objective.

The formulation for time average gap as a linear program is at first nonintuitive. A gap is likely to span multiple time steps, so it is not correct to simply sum the square of the ΔT_{ik} values. Instead, the area under the time to access curve must be calculated. The area for a sample time step is shown in red in Fig. 5.10. The calculation of this area requires the introduction of a reverse accumulator variable, $r_{ik} \in \mathbb{R}_{\geq 0}$. r_{ik} represents the time until the next access at a given time i for station k. The sample area is a trapezoid with a height of ΔT_{ik} (the time step for the area), a base of r_{ik}, and a second base of $r_{ik} - \Delta T_{ik}$. The area is therefore equal to

$$\begin{aligned} \text{area} &= 0.5(r_{ik} + r_{ik} - \Delta T_{ik})\Delta T_{ik} \\ &= r_{ik}\Delta T_{ik} - 0.5\Delta T_{ik}^2 \end{aligned} \tag{5.19}$$

The MTTA can be calculated by summing this equation over all gaps. The TAG is obtained by multiplying the resulting value by two.

The constraints that force r_{ik} to take the proper values are developed. This process is identical to the one used in calculating the accumulator variable for revisit time except that r accumulates backwards in time. If there is access during the current time step, r_{ik} will be

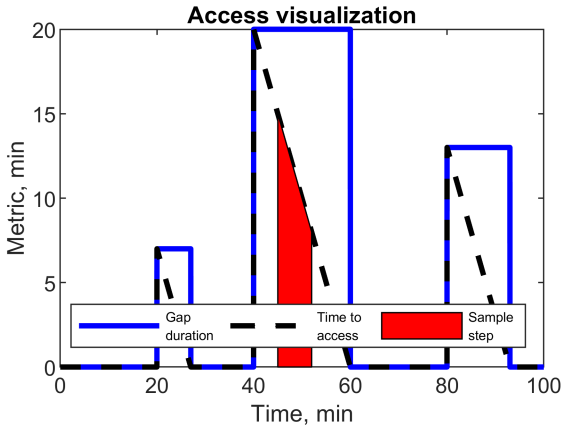


Figure 5.10: Curves for gap duration and time until access.

zero. Otherwise, it will be equal to the time from the beginning of the current time step to the next access. The optimistic assumption is made that if the coverage interval ends on a gap, the gap terminates at the end of the coverage period. The constraints are as follows, where M_2 is as defined in the maximum revisit formulation.

$$r_{(n_t-1)k} \geq \Delta T_{(n_t-1)k} - M_2 Y_{(n_t-1)k} \forall\ k = 1, 2, \ldots, n_g \tag{5.20}$$

$$r_{(n_t-1)k} \leq \Delta T_{(n_t-1)k} + M_2 Y_{(n_t-1)k} \forall\ k = 1, 2, \ldots, n_g \tag{5.21}$$

$$r_{ik} \geq r_{(i+1)k} + \Delta T_{ik} - M_2 Y_{ik} \forall\ i = 1, 2, \ldots, n_t - 2,\ \ k = 1, 2, \ldots, n_g \tag{5.22}$$

$$r_{ik} \leq r_{(i+1)k} + \Delta T_{ik} + M_2 Y_{ik} \forall\ i = 1, 2, \ldots, n_t - 2,\ \ k = 1, 2, \ldots, n_g \tag{5.23}$$

$$r_{ik} \leq M_2(1 - Y_{ik}) \forall\ i = 1, 2, \ldots, n_t - 1,\ \ k = 1, 2, \ldots, n_g \tag{5.24}$$

Eq. (5.19) gives the MTTA when summed over the gaps and divided by the total scenario time (T_{max}). However, while the first term in the equation vanishes for time steps in which there is access, the second term does not. Therefore, multiply the second term by $1 - Y_{ik}$ to ensure that it vanishes when access is present. The resulting equation can be multiplied by two to get TAG, then averaged over the stations to get the average TAG. If instead the maximum or minimum TAG are desired, the process requires the introduction of additional

variables depending on the optimization direction. These formulations are not included but can be developed by combining the process described for average TAG with the process for minimizing or maximizing revisit times described in section 5.2 and Appendix B. The full optimization formulation can be written as

$$\text{Minimize/Maximize } \frac{1}{n_g} \sum_{k=1}^{n_g T_{max}} \sum_{i=1}^{n_t-1} 2r_{ik}\Delta T_{ik} - (1 - Y_{ik})\Delta T_{ik}^2$$

such that:

$$\begin{aligned}
x_j &\in \{0,1\} && \forall\ j = 1,2,\ldots,n_s \\
Y_{ik} &\in \{0,1\} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
r_{ik} &\in \mathbb{R}_{\geq 0} && \forall\ i = 1,2,\ldots,n_t - 1,\ k = 1,2,\ldots,n_g \\
\sum_{j=1}^{n_s} x_j &= n_s - n_r \\
Y_{ik} &\leq \frac{\sum_{j=1}^{n_s} A_{ijk} x_j}{n_c} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
MY_{ik} &\geq \sum_{j=1}^{n_s} A_{ijk} x_j - n_c + 1 && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
r_{(n_t-1)k} &\geq \Delta T_{(n_t-1)k} - M_2 Y_{(n_t-1)k} && \forall\ k = 1,2,\ldots,n_g \\
r_{(n_t-1)k} &\leq \Delta T_{(n_t-1)k} + M_2 Y_{(n_t-1)k} && \forall\ k = 1,2,\ldots,n_g \\
r_{ik} &\geq r_{(i+1)k} + \Delta T_{ik} - M_2 Y_{ik} && \forall\ i = 1,2,\ldots,n_t - 2,\ k = 1,2,\ldots,n_g \\
r_{ik} &\leq r_{(i+1)k} + \Delta T_{ik} + M_2 Y_{ik} && \forall\ i = 1,2,\ldots,n_t - 2,\ k = 1,2,\ldots,n_g \\
r_{ik} &\leq M_2(1 - Y_{ik}) && \forall\ i = 1,2,\ldots,n_t - 1,\ k = 1,2,\ldots,n_g
\end{aligned}$$

This formulation has a total of $n_t \times n_g + n_s$ binary variables and $(n_t - 1) \times n_g$ continuous variables. A similar formulation could be developed for a time average access metric, in which the weighted average of the access durations is taken.

The notation used in the development of this objective is consistent with the notation used in the previous section, which assumes satellites are the viewers providing coverage to ground stations, which are the targets. In the GPS monitoring scenario, the ground stations are viewers providing coverage to the satellites, which are the targets.

5.3.3 Optimization formulation

Mission definition

The existing seventeen monitor stations are assigned as assets to the optimization. The optimizer then sets the coordinates for up to fifteen additional monitor stations. The GPS satellites are used as targets in the coverage analysis. The current GPS configuration consists of six circular orbits, evenly spaced in RAAN, with an inclination of 55° and a semimajor axis of 26,560 km. The analysis is done using the expanded configuration for 27 satellites, with three planes of four satellites and three planes of five satellites [65].

The decision variables for the optimization are shown in Table 5.4.

Table 5.4: Ground station decision variables.

DV	Description	Lower limit	Upper limit	Type
y_{i1}	Ground station latitude (ϕ), °	-90	90	$\mathbb{R}$
y_{i2}	Ground station longitude (λ), °	-180	180	$\mathbb{R}$
$i_{\max}$	Number of additional ground stations	1	15	$\mathbb{Z}$
for i	station ID	1	$i_{\max}$	$\mathbb{Z}$

Optimizer

The following optimizer parameters are used.

- **Optimizer:** Variable-length genetic algorithm with epsilon boxing

- **Initial population:** 400
- **Tournament competitors:** 3
- **Mutation rate:** 5%
- **Length change rate:** 5%
- **Expansion rate:** 0%
- **Nondominated generations for convergence:** 20
- **Number of restarts:** 20
- **Maximum archive size:** 1,000
- **Epsilon boxes:** [0.1, 1, 6.94×10^{-4}, 6.94×10^{-4}, 1]
- **Objective limits:** [Inf, Inf, Inf, Inf, Inf]

Dynamics

The following dynamics parameters are used for the propagation of the satellites.

- **Propagator:** Fast two body propagator
- **Central body:** Earth
- **Scenario epoch:** July 1, 1993, 00:00
- **Scenario duration:** 1 day
- **Propagation time step:** 86.4 sec
- **Satellite pointing:** Aimed

Additionally, the ground station dynamic parameters are:

- **Propagator:** None (position constant in Earth-centered, Earth-fixed frame)
- **Station pointing:** Aimed

Constraints

Two constraints are used in this scenario.

Constraint #1, monitoring stations on land: The GPS monitoring stations can only be constructed on land. Matlab contains a database of coordinates for coastlines. These coordinates are used to determine whether a location is within land boundaries or in the ocean. If δ_L is one if a point is on land and zero otherwise,the constraint equation is:

$$h_{1i}(\mathbf{y}_i) = \delta_L(\mathbf{y}_i) - 1 = 0 \ \forall i = 1, 2, \ldots, i_{\max} \tag{5.25}$$

This constraint is repairable. If the constraint is violated, random combinations of latitude and longitude are generated until a point on land is found. The violating point is replaced with the new point.

Constraint #2, monitoring stations spaced apart: A constraint is imposed that requires that all ground stations be at least 150 km apart. The new ground station positions must be checked against each other and against the 17 existing stations. If $\mathbf{r}_i$ is the position vector for the i-th ground station and $\mathbf{r}_k$ is the position vector for the k-th ground station, the constraint equation is:

$$g_{1i}(\mathbf{y}_i) = \|\mathbf{r}_i - \mathbf{r}_k\| \geq 150 \text{ km } \forall i = 1, 2, \ldots, i_{\max}, \ k = i+1, i+2, \ldots, i_{\max} + 17 \tag{5.26}$$

This constraint is repairable. If the constraint is violated for a pair of ground stations, one ground station position is shifted along the line connecting the pair of ground stations. If this shift puts the station into the water, a random point on land is chosen. This new point replaces the ground station location.

Objectives

This scenario has five objectives.

Objective #1–2, maximize number of assets: Coverage is calculated for each of the 27 GPS satellites in the expanded configuration. The monitoring stations are used as the viewers. The elevation angle of the satellite with respect to the ground station must be at least 10° for access. The ground station is assumed to track all satellites in its field of view simultaneously. The number of ground stations that can access each satellite at each point in time is calculated. The average is taken for each ground station over time. The satellite that has the lowest average is chosen, and its average number of ground stations reported as an objective. This objective is hereafter referred to as the minimum of the average number of assets. Additionally, the minimum number of ground stations available to any satellite at any point in time is found and reported as an objective. This objective is hereafter referred to as the overall minimum number of assets. The epsilon box for the first objective is set at 0.1 and the second objective at 1. If n_{kj} is the number of ground stations that can access satellite k at time step j, Δt_j is the length of time step j, and n_t is the number of time steps, the objective functions are formed by minimizing Eqs. (3.67) and (3.68) across all satellites and multiplying by -1 to transform the maximization problem into a minimization problem:

$$f_1(\mathbf{y}) = -1 \min_k \left(\frac{\sum_{j=1}^{n_t} n_{kj}(\mathbf{y}) \Delta t_j}{\sum_{j=1}^{n_t} \Delta t_j} \right) \tag{5.27}$$

$$f_2(\mathbf{y}) = -1 \min_k \left(\min_j n_{kj}(\mathbf{y}) \right) \tag{5.28}$$

Objective #3–4, minimize degraded average TAG and degraded maximum revisit time: The degraded constellation performance is evaluated assuming a loss of five ground

stations. Degraded performance is evaluated over a period of one day. An MILP solver is used to determine the combination of ground station losses that gives the worst performance with respect to average TAG as described in subsection 5.3.2. The average TAG and the overall maximum revisit time for that degraded solution are used as objectives. The epsilon box for both objectives is set at one minute. If g'_{km} is the length of the m-th degraded gap time for the k-th satellite, n'_{gk} is the number of gaps for the k-th satellite in the degraded scenario, n_s is the number of satellites, and D is the scenario duration, the objective functions are formed by averaging Eq. (3.64) across all satellites and maximizing Eq. (3.62) across all satellites:

$$f_3(\mathbf{y}) = \frac{\sum_{k=1}^{n_s} \sum_{m=1}^{n'_{gk}} (g'_{km}(\mathbf{y}))^2}{n_s D} \tag{5.29}$$

$$f_4(\mathbf{y}) = \max_k \left(\max_m g'_{km}(\mathbf{y}) \right) \tag{5.30}$$

Objective #5, minimize number of ground stations: This objective is set to explore the design space and examine the performance of ground station networks of various sizes. The epsilon box for this objective is set at one ground station. Since there are 17 fixed ground stations, the objective function is

$$f_5(\mathbf{y}) = 17 + i_{\max} \tag{5.31}$$

Mathematical summary

The mathematical formulation for the optimization is shown below.

Minimize

$$
\begin{aligned}
f_1(\mathbf{y}) &= \min_k \left(\frac{\sum_{j=1}^{n_t} n_{kj}(\mathbf{y}) \Delta t_j}{\sum_{j=1}^{n_t} \Delta t_j} \right) \\
f_2(\mathbf{y}) &= \min_k \left(\min_t n_{kj} \right) \\
f_3(\mathbf{y}) &= \frac{\sum_{k=1}^{n_s} \sum_{m=1}^{n'_{gk}} (g'_{km}(\mathbf{y}))^2}{n_s D} \\
f_4(\mathbf{y}) &= \max_k \left(\max_m g'_{km}(\mathbf{y}) \right) \\
f_5(\mathbf{y}) &= 17 + i_{\max}
\end{aligned}
$$

such that

$$
\begin{aligned}
i_{\max} &\in [1, 15] \cap \mathbb{Z} \\
y_{i1} &\in [-90, 90] \ \forall i = 1, 2, \ldots, i_{\max} \\
y_{i2} &\in [-180, 180] \ \forall i = 1, 2, \ldots, i_{\max} \\
g_{1i}(\mathbf{y}_i) &= \|\mathbf{r}_i - \mathbf{r}_k\| \geq 150 \text{ km } \forall i = 1, 2, \ldots, i_{\max}, \ k = i+1, i+2, \ldots, i_{\max} + 17 \\
h_{1i}(\mathbf{y}_i) &= \delta_L(\mathbf{y}_i) - 1 = 0 \ \forall i = 1, 2, \ldots, i_{\max}
\end{aligned}
$$

5.3.4 Results

The simulation produced a nondominated front with 34 solutions. Because the existing monitor stations always provide at least twofold coverage for each satellite, there is no benefit in studying revisit metrics such as maximum revisit time for the nominal case. Instead, the minimum of the average number of assets and the overall minimum number of assets for each solution are shown in Fig. 5.11. The points that are nondominated (ND) in number of ground stations and the minimum of the average number of assets are marked with a

circle, points that are nondominated in number of ground stations and the overall minimum number of assets are marked with a triangle, and points that are dominated (D) in number of ground stations and assets (but nondominated across all five objectives) are marked with an x. As would be expected, the minimum of the average number of ground stations available increases linearly with the total number of ground stations. Because this objective has no upper bound, it is beneficial to add as many ground stations as possible. Thus, some solutions have the full fifteen additional ground stations allowed in the optimization. Conversely, the objective to minimize the number of ground stations drives some other solutions toward the minimum number of additional ground stations, which is one. The minimum number of ground stations available varies among solutions, with some of the larger networks having at least six ground stations in view of every satellite at all times.

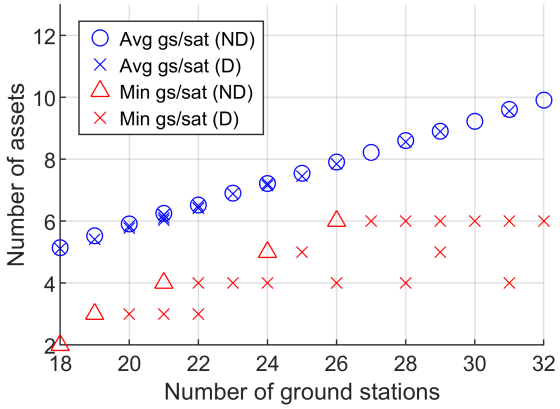


Figure 5.11: Nondominated front for ground station expansion, number of assets.

Fig. 5.12 shows the degraded average TAG and the degraded maximum revisit time for each solution when five ground stations have been lost. The x axis shows the number of ground stations in the nominal (non-degraded) solutions. The points that are nondominated (ND) in number of ground stations and the degraded average TAG are marked with a triangle, points that are nondominated in number of ground stations and the degraded maximum revisit time are marked with an inverted triangle, and points that are dominated (D) in

number of ground stations and degraded revisit metrics (but nondominated across all five objectives) are marked with an x. For the case in which only a single ground station has been added to the seventeen existing stations, the degraded maximum revisit time for at least one of the satellites is between 3 and 3.5 hours. Because that satellite's orbit will not be monitored for over three hours, corrections to the orbit cannot be made and accuracy will degrade. These solutions highlight the vulnerabilities in the existing monitor station network. By contrast, high numbers of ground stations provide continuous monitoring even when five stations are disabled. The addition of nine ground stations allows for continuous monitoring in the degraded case.

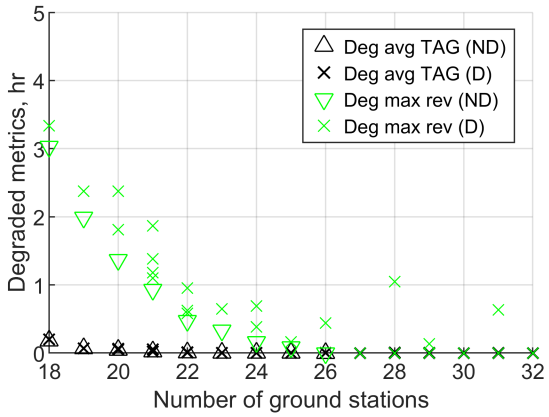


Figure 5.12: Nondominated front for ground station expansion, degraded revisit metrics.

The placement of ground stations for this 26 station case is shown in Fig. 5.13. Three stations are placed in Antarctica due to the lack of existing stations below 45°S. These stations are geographically close to one another but provide the redundancy necessary to achieve continuous coverage in the degraded case. If greater positional diversity is desired, the minimum distance between each station can be increased. A ground station is placed to the far north in Russia, as the existing network has only two stations above 45°N. Three more stations are placed near the equator in Africa and South America. One station each is placed in India and Oceania.

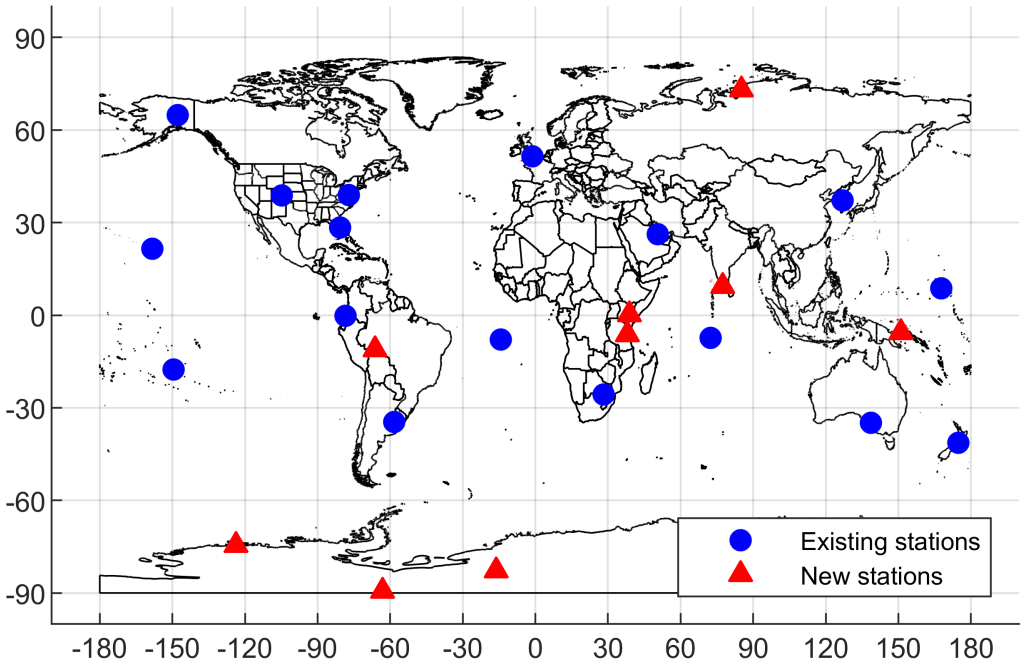


Figure 5.13: Expanded 28 station network.

Fig. 5.14 shows the ground stations added across all 34 solutions. By looking at all solutions simultaneously, it is clear which regions most frequently hosted additional ground stations. Antarctica, Russia, and Greenland frequently had ground stations assigned due to the lack of coverage in polar regions. Africa had many stations assigned, as did India. This suggests a need for increased coverage in the Indian Ocean region and over the equatorial region of Africa. Several solutions place stations in Brazil and neighboring countries, suggesting a need for additional infrastructure in that region. Finally, the eastern coast of Australia and Melanesia had occasional additions. These regions were likely favored due to the lack of large land masses in the Pacific Ocean. Since points must either be bred from existing solutions or randomly selected, the probability of selecting a point on a small island in the Pacific is unlikely. Therefore, points on the nearest large land masses were chosen instead. The likelihood of placing ground stations on islands could be increased by modifying the repair

operation so it chooses the nearest land point rather than a random land point. Alternately, a penalty function could be used that penalizes distance from land.

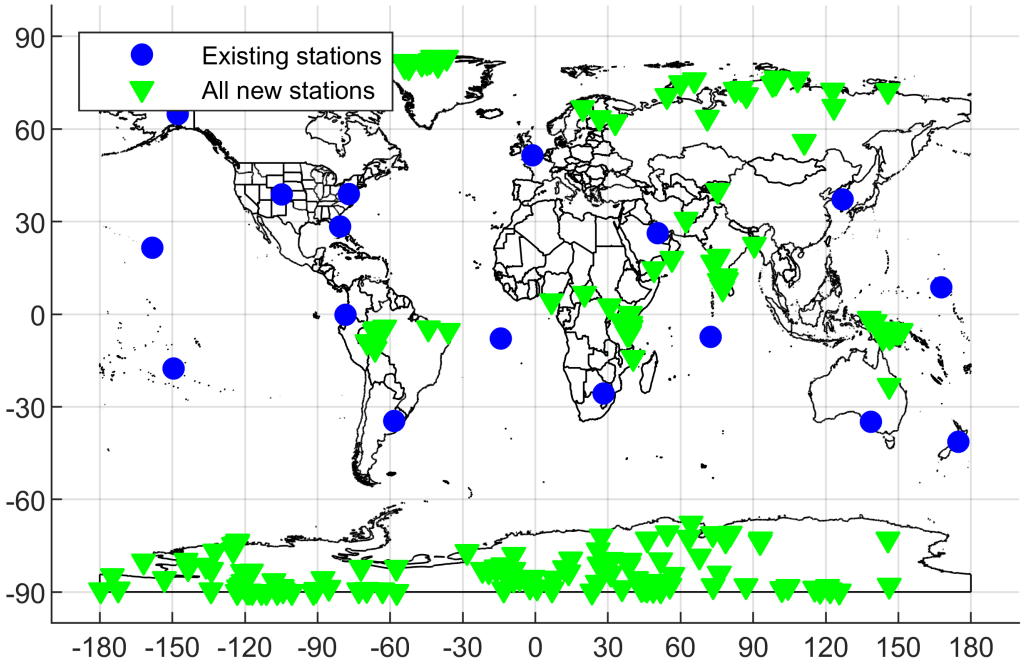


Figure 5.14: Ground stations across all solutions.

This initial analysis does not account for limitations on monitor station placement, such as the difficulty in maintaining stations in remote regions or infeasible locations due to political boundaries. However, the formulation can easily be modified to exclude certain regions. As a demonstration, the analysis was rerun with the constraint that the monitor stations must only be placed in a country that is part of the Five Eyes (FVEY) alliance: Australia, Canada, New Zealand, the United Kingdom, and the United States. Territories of the United States were also considered permissible. The repair method for the "monitoring stations on land" constraint was also modified to shift an inadmissible point to the nearest admissible location rather than randomly selecting a new location.

This new analysis produced a nondominated front containing 97 solutions. The minimum of

the average number of assets and the overall minimum number of assets for each solution are shown in Fig. 5.15. The solutions produced by this analysis generally have a lower overall minimum number of assets than in the original analysis due to the lack of positional diversity, particularly at the poles. Fig. 5.16 shows the degraded average TAG and the degraded maximum revisit time for each solution when five ground stations have been lost. The x axis shows the number of ground stations in the nominal (non-degraded) solutions. Because of the newly imposed limitation on the permissible station locations, no solutions were found that provided continuous coverage in the degraded case. However, degraded performance can be improved with the addition of new stations. For example, five additional stations can decrease the degraded average TAG from greater than 16 minutes to 7 minutes or decrease the degraded maximum revisit time from greater than 3.5 hours to 2.5 hours.

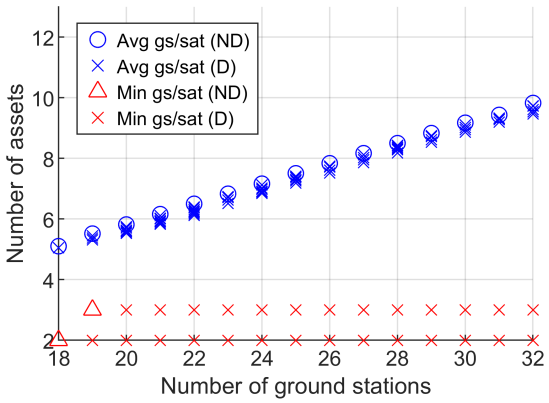


Figure 5.15: Nondominated front for FVEY ground station expansion, number of assets.

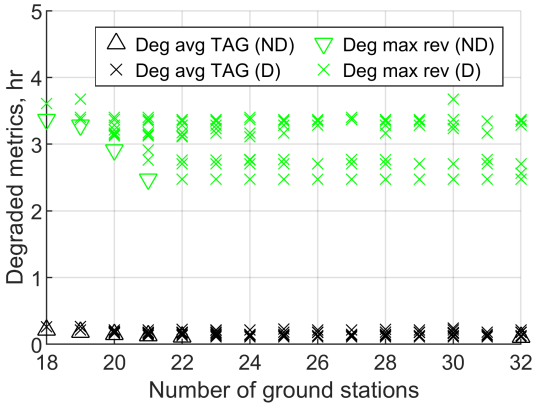


Figure 5.16: Nondominated front for ground station expansion, degraded revisit metrics.

The placement of new stations for one potential solution is shown in Fig. 5.17. This solution has five additional ground stations. The minimum of the average number of assets is 6.1, and the overall minimum number of assets is 3. The degraded average TAG is 6.7 minutes, and the degraded maximum revisit time is 2.7 hours.

Fig. 5.18 shows the ground stations added across all 97 solutions. Northern Canada fre-

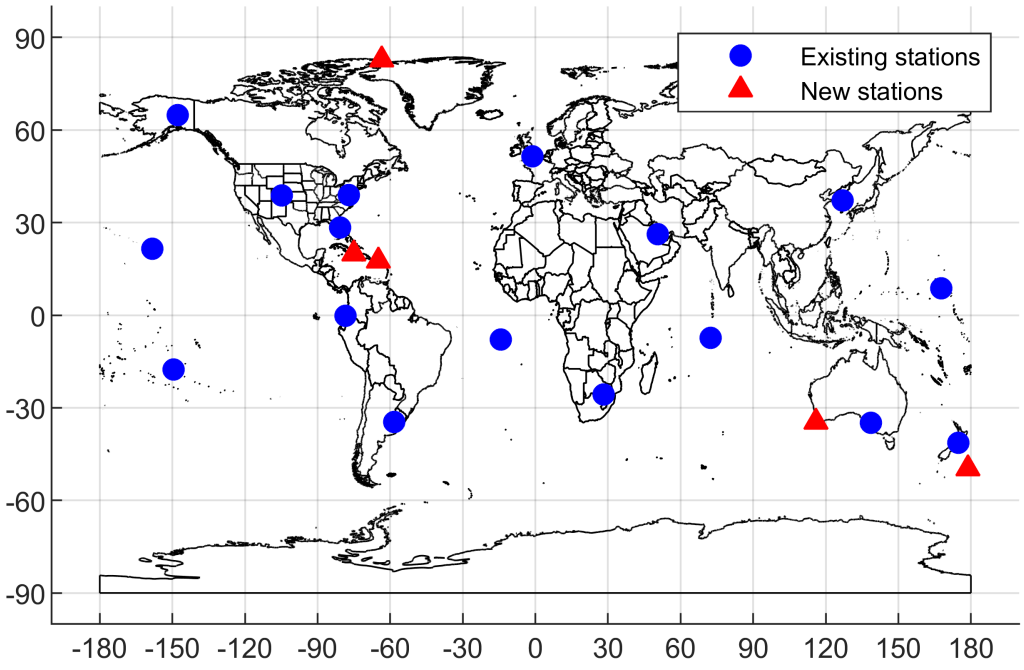


Figure 5.17: Expanded FVEY station network.

quently had ground stations assigned due to the lack of coverage in polar regions. Western Australia also had many stations assigned due to a need for increased coverage in the Indian Ocean region. There were several additions in various parts of Oceania, such as New Zealand and Guam, to increase coverage in the Pacific Ocean. Finally, several solutions place stations in the Caribbean and in the United Kingdom, suggesting that additional infrastructure in these regions would be beneficial.

These analyses do not account for the geographic distribution of user needs. For many applications, a loss of accuracy in the polar regions would not degrade operational capability. Additional studies could restrict analysis to times when the satellites are over regions of interest, such as particular cities or war zones. Other variations could restrict the ground stations to other regions or could optimize the network based on a predefined list of potential stations.

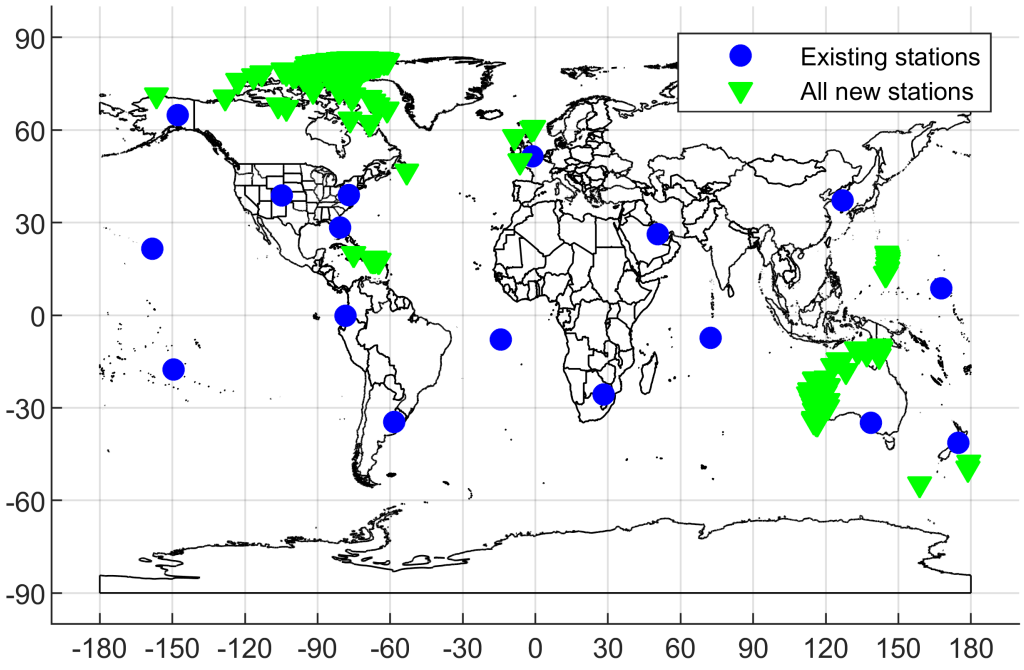


Figure 5.18: FVEY ground stations across all solutions.

When considering the trade-offs between cost, performance, resilience, and complexity for this scenario, the number of stations can be used as a surrogate for cost. Increasing the cost therefore increases the resilience of the network. However, because the nominal case provides sufficient coverage with the existing seventeen stations, increasing cost does not increase nominal performance according to these analysis. More detailed analyses that account for sensor accuracy may see an increase in the nominal performance due to the synthesis of information from multiple sources. Complexity is not considered in this scenario, as it is assumed that the configuration of an individual ground station is independent of the existence of other ground stations.

5.3.5 Conclusions

This scenario describes an extension of the previously discussed resilience methodology that uses average time average gap as the objective that determines performance. The number of assets was examined for the nominal case, and the average TAG and maximum revisit times were examined in the degraded case where four ground stations were unavailable. It also demonstrates the applicability of the resilience methodology and the larger DISCO-Tech methodology to the optimization of ground station networks.

Chapter 6

Applications: Trade-offs for ocean altimetry

6.1 Introduction

Disaggregation offers unique benefits and challenges compared to monolithic systems. These benefits and challenges are exemplified by considering a set of payloads with fixed design parameters. If these payloads are placed on a single monolithic system, the resulting satellite could be expected to have the lowest possible mass and cost. A single satellite would simplify ground operations. Additionally, a single satellite eliminates the need for any crosslinking or formation flying. However, monolithic systems offer no resilience to threat or failure, as the loss of the single satellite destroys all capability to perform the mission. The individual satellite mass is comparatively large. Finally, conflicting orbital needs by various payloads may force the satellite into a sub-optimal orbit for every mission. Comparatively, disaggregating the payloads across one or more satellites increases the resilience of the system, particularly if redundant payloads are included. Because each satellite has less than the full suite of payloads, the individual mass of each satellite is less than in the monolithic case, potentially reducing the complexity of the individual satellites. Additionally, lower individual mass may increase the number of available launch options. Finally, different orbits can be selected for each satellite based on the needs of that satellite's payloads. However, the use of multiple

satellites increases the total cost and mass of the system and requires more complex ground operations. Depending on payload needs, crosslinking or formation flying may be necessary as well.

Complexity is a driver for disaggregation. Government-sponsored satellite missions often suffer from schedule delays and cost overruns, largely due to the complexity of monolithic satellites and the fact that many satellites require unique and never-before flown payloads and hardware. In response to these issues, there has been a growing trend in small, low-cost satellites in recent years [137]. If mission payloads are still under development, it may be possible to use simplified payloads across multiple satellites to accomplish the mission. In addition to potential cost benefits, these disaggregated architectures decrease mission risk both by lowering the complexity of the constituent satellites and by increasing redundancy and resiliency in the event of a single spacecraft anomaly. However, systems employing simplified payloads may suffer from performance loss compared to monolithic systems.

To select the appropriate level of disaggregation for a mission, it is necessary to quantify the advantages and disadvantages of different configurations and the trade-offs between objectives. This chapter addresses the third research question: What are the trade-offs between cost, performance, resilience, and complexity for disaggregated and non-traditional systems?

6.2 Scenario 6-a: Heterogeneous InSAR formation for ocean topography

6.2.1 Purpose and background

One field of study that may benefit from disaggregation is climate monitoring. Sea surface height (SSH) and surface water levels are two key parameters in global climatology and resource management. National interest in the expansion of the measurement of water levels for both surface and ocean water was highlighted in the 2007 National Research Council Decadal Survey [32]. The analyses of weather monitoring disaggregation in Refs. 151 and 95 cover some aspects of climate monitoring, but not water level monitoring. Ocean currents affect SSH, so tracking SSH allows scientists to predict weather patterns like El Niño in addition to tracking long-term changes in ocean level. SSH has been studied using satellite remote systems in the past, starting with the joint NASA/National Centre for Space Studies (CNES) TOPEX/Poseidon mission that was in operation from 1992 to 2006. TOPEX/Poseidon used a dual-frequency radar altimeter to map SSH on a ten-day repeating ground track that covered 95% of ice-free oceans. These measurements had an accuracy of better than five centimeters, while the system had knowledge of its own position within ten centimeters [1]. TOPEX/Poseidon was succeeded by the Jason series of satellites when Jason-1 launched in 2001 and operated until 2013. Jason-2 and Jason-3, which launched in 2008 and 2016 respectively, are currently in operation. The Jason satellites use a similar suite of instruments to TOPEX/Poseidon, including a Poseidon dual-frequency radar altimeter. The next iteration of Jason will be incorporated into the Sentinel mission as Sentinel-6/Jason-CS in 2020 [130]. Although nadir altimeters offer high accuracy measurements, their lack of wide-swath coverage makes it impossible to measure sub-mesoscale processes, where "mesoscale" refers

to phenomenon with spatial scales between 50 and 500 km and temporal scales between 10 and 100 days.

The NASA/CNES Surface Water and Ocean Topography (SWOT) mission aims to make these sub-mesoscale measurements as well as measurements of surface water height through the use of interferometric synthetic aperture radar (InSAR). However, like many complex and novel spacecraft projects, SWOT is experiencing both cost and schedule challenges. According to a 2018 audit, critical design reviews were delayed for eight of eleven payload modules, and projections predict that the project may not meet its launch goal of April 2021. The estimated life cycle cost for the program is $1.14 billion, but technical issues in the primary payload and cost modelling deficiencies led the audit to suggest that the allocated funds may be insufficient [5]. This analysis examines whether the goals of the SWOT program can be met using a disaggregated space system to perform wide-swath InSAR. The analysis focuses on SSH monitoring, but the methodology is designed for flexibility and can be extended to surface water monitoring.

InSAR requires the comparison of two or more synthetic aperture radar (SAR) images of the same target location. These images must be acquired using receivers at different points in space. For the purpose of height detection using InSAR, the receivers must be separated in the cross-track direction when the images under comparison are captured. There are two distinct methods for collecting these images: simultaneous interferometry in which the images are collected using two different receivers at the same point in time, and multi-pass interferometry in which the images are are collected at different points in time by a single spacecraft [102]. Missions using multi-pass interferometry with a single satellite include ERS-1 and 2 [84], Envisat [111], TerraSAR-X [21], and the upcoming NISAR mission [170]. Alternately, separate satellites flying over the same ground points at different times can be used to generate the SAR images. An example of this type of mission is the ERS-1/2 tandem

demonstration.

Simultaneous interferometry can either be performed by a single satellite with multiple on-board receivers or by multiple satellites flying in formation. An example of a single-vehicle simultaneous interferometry mission is the NASA Shuttle Radar Topography Mission (SRTM), which was flown for eleven days in 2000 to model land surface elevation [52]. SWOT will also be a single-vehicle simultaneous interferometry mission. Multi-satellite InSAR systems are called Distributed Satellite-Borne SAR (DSS). DSS systems can either have a transmitter-receiver pair on every satellite (full-active) or have some satellites with transmitter-receiver pairs and some satellites with receivers only (semi-active) [164]. For DSS missions, precise knowledge of the baseline (the distance between the receivers) is critical and greatly impacts the error of the generated height measurements. The first DSS system was the TerraSAR-X/TanDEM-X mission, where the two satellites alternated illumination of a ground area simultaneously monitored by both satellites [21]. Simultaneous interferometry has the advantage of not suffering from temporal decorrelation.

Despite the established history of satellite-based InSAR, there are few modelling resources for multi-satellite InSAR systems. The main contribution of this work is a methodology for modelling and optimizing cost and performance for disaggregated InSAR systems. This analysis will examine the performance of functionally disaggregated systems by simultaneously optimizing payload parameters, payload distribution, and orbital parameters and will use receivers placed on different satellites to form the baseline for the system to determine if simplifying the payload can reduce cost.

6.2.2 Methodology

This analysis uses the DISCO-Tech methodology with custom objectives and constraints. The mission requirements and parameters are based largely on those dictated by the SWOT program. SWOT's primary payload has several planned components [3]:

- Ka-band Radar Interferometer (KaRIn): KaRIn consists of two Ka-band transmit-receive units separated by a 10 m baseline. The transmitters have two different polarizations (VV and HH) and illuminate swaths on separate sides of the satellite path.
- Jason-class nadir altimeter: This radar altimeter will make measurements of SSH directly below the satellite, providing data in the gap uncovered by KaRIn.
- Microwave radiometer: The microwave radiometer, an extension of the Advanced Microwave Radiometer (AMR) flown on the Jason satellites, collects data on water vapor in the troposphere and is used to remove errors induced by signal delay due to the water vapor.
- X-band data downlink: SWOT requires a high data rate and plans to use an X-band downlink with the CNES X-band ground station network.
- Precision orbit determination hardware: SWOT will include a GPS receiver, a DORIS receiver, and a laser retroreflector array for precision orbit determination.

Formation flying

The formation used by the spacecraft must be defined. A small eccentricity difference δ_e with respect to a circular reference orbit will cause a satellite to move about an elliptical relative orbit in the radial/in-track plane around the nominal trajectory. Fig. 6.1 shows this relative

in-track/radial trajectory for three satellites flying in formation around a circular reference orbit. The satellites are spaced in argument of periapsis. Additionally, RAAN/inclination differences ($\Delta\Omega$ and Δi) in the presence of a slight eccentricity (e) will result in relative motion in the radial/cross-track plane [37]. Fig. 6.2 shows the relative cross-track/radial trajectory for a single satellite flying in formation around a circular reference orbit, where the cross-track motion is governed by a small change in inclination δ_i. However, the relative phase between the radial and cross-track motion is a function of the change in inclination, change in RAAN, argument of periapsis, and eccentricity. This relative phase dictates the orientation and flattening of the ellipse. The general form of the relative cross-track/radial trajectory is shown in Fig. 6.3. Additionally, a change in RAAN shifts the in-track/radial relative motion ellipse along the in-track axis by a constant amount. This consequence is not considered in the linearization in Ref. 37 but can be corrected by offsetting the initial mean arguments of latitude of the satellites and so is not discussed further in this analysis.

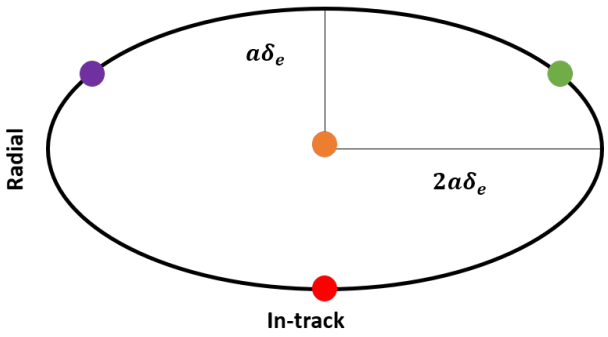


Figure 6.1: Relative motion in the in-track/ radial plane for a small eccentricity and varying ω.

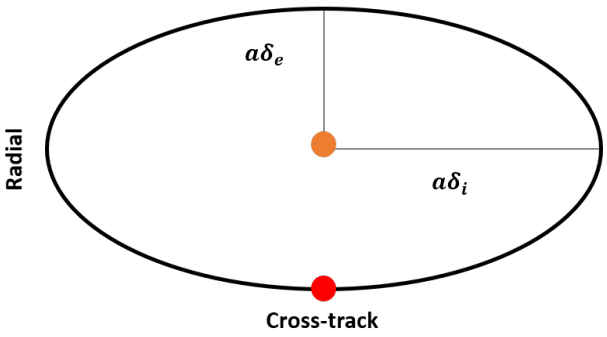


Figure 6.2: Relative motion in the cross-track/radial plane for a small eccentricity, $\omega = 0$, and small inclination change.

This formation is called a helix formation. In the case of a multi-satellite formation, it is desirable to have all of the satellites follow the same relative path to establish a consistent baseline. Relationships between the argument of periapsis, RAAN, and inclination of the

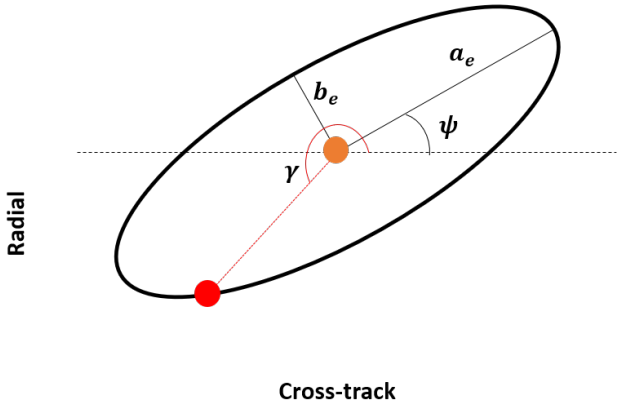


Figure 6.3: Relative motion in the cross-track/radial plane, general.

satellites must be found that provide motion along the same path. The orbital elements of the circular reference orbit are the semimajor axis a, the inclination i, the RAAN Ω, and the mean argument of latitude u_r. Because the separation in the radial/cross-track plane determines the baseline for height measurements, the relative path in these coordinates is the focus of the derivation. If there is no difference in a relative to the circular reference orbit, the linearized relative equations of motion in the radial and cross-track directions are

$$\Delta r_R/a = -\Delta e_x \cos(u) - \Delta e_y \sin(u) \tag{6.1}$$

$$\Delta r_N/a = -\Delta i_y \cos(u) + \Delta i_x \sin(u) \tag{6.2}$$

$$\Delta i_x = \Delta i = \delta_i \cos(\theta) \tag{6.3}$$

$$\Delta i_y = \Delta\Omega \sin(i) = \delta_i \sin(\theta) \tag{6.4}$$

$$\Delta e_x = e\cos(\phi) = \delta_e \cos(\phi) \tag{6.5}$$

$$\Delta e_y = e\sin(\phi) = \delta_e \sin(\phi) \tag{6.6}$$

where u is the mean argument of latitude and ϕ and θ are angular parameters that dictate the satellite's initial position on the relative orbit. Because the reference orbit is circular, ϕ

is equal to the argument of periapsis of the satellite. For the satellites to follow the same path, it must be true that the relative position of one satellite at some mean argument of latitude u_1 will be equal to the position of another satellite at some mean argument of latitude $u_2 = u_1 + \delta_u$. Each satellite in the formation will have a unique value for θ and ϕ called θ_j and ϕ_j. Using Eqs. (6.1), (6.5), and (6.6) and the knowledge that the satellites must follow the same relative path, it can be derived that $\delta_i \sin(u_1 - \theta_1) = \delta_i \sin(u_1 + \delta_u - \theta_2)$. Likewise from Eqs. (6.2), (6.3), and (6.4), $-\delta_e \cos(u_1 - \phi_1) = -\delta_e \cos(u_1 + \delta_u - \phi_2)$. Because these relationships must hold for any value of $u_1 \in [0, 2\pi]$, it must be true that $\phi_2 = \delta_u + \phi_1$ and $\theta_2 = \delta_u + \theta_1$.

To evenly space n_s satellites around the path, the relative argument of latitude $\delta_{u,j}$ for the j-th satellite is

$$\delta_{u,j} = \frac{2\pi(j-1)}{n_s} \tag{6.7}$$

The satellites will move along an ellipse in the radial/cross-track plane. The ellipse has a semimajor axis a_e, a semiminor axis b_e, and a rotation angle about the in-track axis of ψ. The general equation for an ellipse is

$$\left(\frac{\cos^2\psi}{a_e^2} + \frac{\sin^2\psi}{b_e^2}\right)x^2 + 2\cos\psi\sin\psi\left(\frac{1}{a^2} - \frac{1}{b^2}\right)xy + \left(\frac{\sin^2\psi}{a_e^2} + \frac{\cos^2\psi}{b_e^2}\right)y^2 = 1$$

The Cartesian coordinates are $x = \Delta r_N$ and $y = \Delta r_R$, and the circular coordinates are $r^2 = x^2 + y^2$ and $\gamma = \tan^{-1}(y/x)$. By definition, r is maximized when $\gamma = \psi$ and is equal to the relative semimajor axis a_e. Similarly, r is minimized when $\gamma = \psi + \pi/2$ and is equal to the relative semiminor axis b_e. Since $x = r\cos(\gamma)$ and $y = r\sin(\gamma)$, the following equations can be formed from Eqs. (6.1)–(6.6):

$$a\delta_i \sin(u_{1\psi} - \theta_1) = a_e \cos(\psi)$$

$$-a\delta_e \cos(u_{1\psi} - \phi_1) = a_e \sin(\psi)$$
$$a\delta_i \sin(u_{1\psi} + \pi/2 - \theta_1) = b_e \cos(\psi + \pi/2)$$
$$-a\delta_e \cos(u_{1\psi} + \pi/2 - \phi_1) = b_e \sin(\psi + \pi/2)$$

where a denotes the orbital semimajor axis and $u_{1\psi}$ is the mean argument of latitude when the satellite is at the maximal distance in the radial/cross-track plane from the reference orbit. Combining these equations forms relationships between δ_e, δ_i, θ_1, ϕ_1, and the ellipse parameters,

$$\delta_i = \frac{1}{a}\sqrt{a_e^2 \cos^2(\psi) + b_e^2 \sin^2(\psi)} \tag{6.8}$$
$$\delta_e = \frac{1}{a}\sqrt{a_e^2 \sin^2(\psi) + b_e^2 \cos^2(\psi)} \tag{6.9}$$
$$\theta_1 - \phi_1 = \sin^{-1}\left(\frac{b_e \cos(\psi)}{a\delta_e}\right) - \sin^{-1}\left(\frac{a_e \cos(\psi)}{a\delta_i}\right) \tag{6.10}$$

The selection of ϕ_1 dictates the initial position of the first satellite along the elliptical relative path. It is equal to the argument of periapsis of the satellite.

Once ϕ_1 is selected, there is sufficient information to calculate the orbital element set for each satellite from the relative ellipse parameters using Eqs. (6.3)–(6.10). Additionally, the satellites should have the same mean argument of latitude to ensure they move together. If the subscript r denotes the reference orbit, the orbital elements for the j-th satellite are

$$a_j = a$$
$$e_j = \delta_e$$
$$i_j = i + \delta_i \cos(\theta_j)$$
$$\omega_j = \phi_j$$

$$\Omega_j = \Omega + \delta_i \frac{\sin(\theta_j)}{\sin(i_r)}$$

$$M_j = u_r - \phi_j$$

where ω is the argument of periapsis and M is the mean anomaly.

Sea surface height error

The height measurement error is a key metric for satellite altimetry. A variety of sources contribute to height error. Key sources are systematic errors due to baseline and angle uncertainty, error due to absolute position uncertainty, and phase errors due to decorrelation. These errors can be converted to height errors using geometric relationships. For this analysis, the geometry of the InSAR system is dictated by the relative motion of the satellites. A representative trajectory for two satellites is shown in Fig. 6.4. The geometry of an InSAR system is shown in Fig. 6.5.

It is first assumed that the interferometric image is formed using one monostatic and one bistatic radar. The satellite at point A_1 transmits, the signal is reflected at the target, and the reflected signal is received at the monostatic radar at point A_1 and the bistatic radar at point A_2. The phase of the signal measured at A_1 (after unwrapping) is $\phi_1 = 2\pi/\lambda(r+r)+\epsilon_1$, where λ is the wavelength. The phase of the signal measured at A_2 is $\phi_2 = 2\pi/\lambda(r+r_2)+\epsilon_2$. If the measurements are sufficiently correlated such that the residuals ϵ_1 and ϵ_2 are approximately equal, the phase difference between the two measurements is $\Delta\phi = -2\pi/\lambda(r_2 - r)$. If instead the interferogram is formed using two monostatic radars, the phase difference is $\Delta\phi = -4\pi/\lambda(r_2 - r)$. Therefore, the phase difference can generally be written as

$$\Delta\phi = -pk(r_2 - r) \tag{6.11}$$

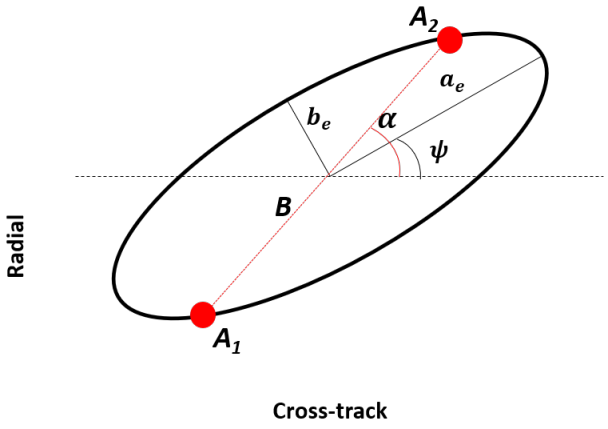


Figure 6.4: Geometry in the cross-track/radial plane for two satellites relative to a circular reference orbit.

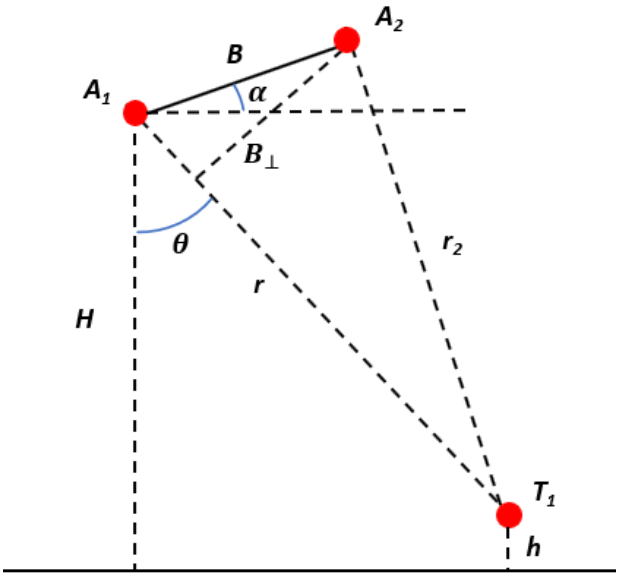


Figure 6.5: Basic InSAR geometry, radial/ cross-track plane.

where k is the wavenumber and p is one if the receivers use the same transmitter and two if each receiver uses its own collocated transmitter. The difference in path length Δ_r is defined such that $r_2 = r + \Delta_r$. From the law of cosines,

$$(r + \Delta_r)^2 = r^2 + B^2 - 2rB\cos\left(\frac{\pi}{2} - \theta + \alpha\right)$$

where B is the baseline, θ is the look angle, and α is the roll angle. Rearranging yields

$$\begin{aligned}\frac{(r + \Delta_r)^2 - r^2 - B^2}{2rB} &= -\cos\left(\frac{\pi}{2} - \theta + \alpha\right) \\ &= -\sin(\theta - \alpha)\end{aligned}$$

Both the baseline and the difference in path length are generally much smaller than the path length, so it can be assumed that $\Delta_r^2, B^2 << r\Delta_r$. The above equation then becomes

$-\sin(\theta-\alpha) \approx \Delta_r/B$. Eq. (6.11) is therefore

$$\Delta\phi = pkB\sin(\theta-\alpha) \tag{6.12}$$

Since the baseline, wavenumber, and roll angle are known quantities, the phase difference is used to obtain the look angle of the received signal. The target's surface height h with respect to some reference surface can be calculated as

$$h = H - r\cos(\theta) \tag{6.13}$$

where H is the satellite's height above the reference surface, which is assumed to be known perfectly. The relationship between phase errors and height errors must be found. Phase errors manifest as errors in the measured look angle, so the height error due to the phase error is

$$\delta_h = \frac{\partial h}{\partial r}\delta_r + \frac{\partial h}{\partial\theta}\frac{\partial\theta}{\partial\Delta\phi}\delta_{\phi,tot} \tag{6.14}$$

Obtaining the partial derivatives from Eqs. (6.12) and (6.13), the height error is

$$\delta_h = -\cos(\theta)\delta_r + \frac{r\sin(\theta)}{pkB\cos(\theta-\alpha)}\delta_{\phi,tot} \tag{6.15}$$

δ_r is the absolute position error and $\delta_{\phi,tot}$ is the phase error from all sources. The phase error can be further split into contributions from decorrelation (δ_ϕ), baseline error (δ_B), and roll error (δ_α), so $\delta_{\phi,tot} = \delta_\phi + (\partial\Delta\phi/\partial B)\delta_B + (\partial\Delta\phi/\partial\alpha)\delta_\alpha$. The partial derivative terms can be found from Eq. (6.12) as $\partial\Delta\phi/\partial B = pk\sin(\theta-\alpha)$ and $\partial\Delta\phi/\partial\alpha = -pkB\cos(\theta-\alpha)$. The total height error is therefore

$$\delta_h = -\cos(\theta)\delta_r + \frac{r\sin(\theta)\Big(\delta_\phi + pk\sin(\theta-\alpha)\delta_B - pkB\cos(\theta-\alpha)\delta_\alpha\Big)}{pkB\cos(\theta-\alpha)} \tag{6.16}$$

For receivers on two different satellites, both the baseline and roll errors are a function of the relative position error and cannot be treated independently. It is assumed that the phase error and baseline/roll errors are uncorrelated. The statistical height error is therefore

$$\sigma_h^2 = \cos^2(\theta)\sigma_r^2 + \frac{r^2 \sin^2(\theta)\left(\sigma_\phi^2 + (pk\sin(\theta-\alpha)\sigma_B - pkB\cos(\theta-\alpha)\sigma_\alpha)^2\right)}{p^2k^2B^2\cos^2(\theta-\alpha)} + \sigma_o^2 \qquad (6.17)$$

where σ_o represents the error due to any other source. The σ terms denote the statistical error and replace the deterministic δ terms in Eq. (6.16).

The magnitude of the various error sources will now be discussed. For a disaggregated interferometric formation, the baseline is formed by the distance between the two receivers and is therefore not physically constrained as it would be in a system with collocated receivers like SWOT. The baseline and angle errors are therefore determined by the relative position uncertainty between the two spacecraft. The GRACE and TerraSAR-X/TanDEM-X missions were able to show that relative position errors on the order of one millimeter were achievable through the use of differential and double differential GPS [90, 175]. GRACE had a baseline between 170 km and 270 km, larger than any considered for the interferometric SAR system, while TerraSAR-X/TanDEM-X had a comparatively small separation of a couple of kilometers. The one millimeter error experienced by these systems can therefore be used to reasonably bound the achievable performance for relative positioning accuracy of a distributed InSAR system, so a baseline error σ_B of 1 mm is used in the analysis. σ_α is derived from σ_B using geometry.

The error due to absolute position uncertainty maps directly to an equivalent absolute height error. SWOT predicts an absolute position error of only 0.14 cm due to the inclusion of its precision orbit determination system [4]. The DORIS receiver alone gives an absolute

position error of 5 cm, and a TriG GPS receiver can also be expected to give centimeter-level accuracy [17, 80]. If a satellite is assumed to have a comparable POD system, the absolute position error can be ignored due to its low magnitude. Additionally, systematic errors in absolute position will not affect the relative height measurements between nearby points, only absolute measurements. If at least one of the satellites has low absolute position error through an on-board orbit determination system, differential GPS can be used to obtain low absolute position error for nearby satellites as well. For these reasons, absolute position uncertainty is not considered in this analysis, so σ_r is assumed to be zero.

Many uncertainties manifest as phase errors. Some potential sources of phase error are volumetric decorrelation due to the presence of multiple scatterers in a single resolution cell, baseline decorrelation, temporal decorrelation, thermal/noise decorrelation due to a finite signal-to-noise ratio (SNR), and decorrelation due to misregistration. Volumetric decorrelation due to ocean waves can be estimated as [140]

$$\gamma_V \approx \exp\left[-\frac{(kB\sigma_s\cos(\theta-\alpha))^2}{2(r\sin\theta)^2}\right] \tag{6.18}$$

σ_s is the wave height standard deviation and is one quarter of the significant wave height. Since SWOT defines significant wave height as 2 m, the analysis uses $\sigma_s = 0.5$ m. Baseline correlation approaches zero as the perpendicular baseline approaches the critical baseline [58]:

$$\gamma_B = 1 - \frac{B\cos(\theta-\alpha)}{B_c} \tag{6.19}$$

where the critical baseline is

$$B_c = \frac{2r\lambda B_w\tan(\theta)}{pc} \tag{6.20}$$

B_w is the transmitter bandwidth.

Temporal decorrelation occurs when the two SAR images used to generate the interferogram are collected at different times. Images are collected simultaneously, so temporal decorrelation is not a concern. For mesoscale measurements, the common decorrelation time used is ten days. Simulations suggest that for sub-mesoscale processes on the order of 30 km, the decorrelation coefficient is about a day [112, 153]. Decorrelation due to misregistration is also neglected in the current analysis.

Thermal decorrelation is a function of the SNR [140],

$$\gamma_N = \frac{\text{SNR}}{\text{SNR} + 1} \tag{6.21}$$

The SNR for a SAR system is [36]

$$\text{SNR} = \frac{\sigma_0 P_t G_t G_r \lambda^3 c\tau \text{PRF}}{4^4 \pi^3 r^3 v \sin(\theta - \alpha) k_B T B_w F L} \tag{6.22}$$

where σ_0 is the normalized backscattering coefficient, P_t is the transmit power in watts, G_t and G_r are the transmit and receive antenna gains in W/W, c is the speed of light in m/s, k_B is Boltzmann's constant, T is the reference temperature in kelvin, F is the noise figure, and L represents the various losses. The loss term is based on the loss of 4.1 dB experienced by TanDEM-X plus atmospheric loss [94]. The noise figure is estimated at 4.3 dB, again based on TanDEM-X [94].

The normalized backscattering coefficient is taken from the SWOT documentation as [4]

$$\sigma_0(w) = 0.52 q(w) e^{-q(w) \tan^2 \theta} \tag{6.23}$$

where

$$q(w) = \frac{1}{0.019\log\left(8.35\sqrt{-\log(1-w)}\right)} \tag{6.24}$$

The wind speed percentile w is set at 0.68. This formulation does not account for the impact of frequency or polarization on backscattering. The equation was derived for a Ka-band radar but notes similar performance in the Ku band. Previous studies suggest that ocean backscatter in the Ku and C bands are similar under high wind conditions, so for this study the same backscattering equation is used for all frequencies [44].

The three decorrelation values stated above are combined to obtain the total decorrelation,

$$\gamma_T = \gamma_V \gamma_B \gamma_N \tag{6.25}$$

The phase error can be estimated using the Cramer-Rao bound (valid if $N > 4$):

$$\sigma_\phi = \sqrt{\frac{1-\gamma_T^2}{2N\gamma_T^2}} \tag{6.26}$$

where N is the number of independent samples over which the measurement is averaged [58]. The number of samples used in the averaging, also called the number of looks, can be obtained by calculating the number of resolution cells that fit within the desired output resolution. For SAR, the azimuthal (in-track) resolution is $d_a/2$, and the ground resolution is $c/(2B_w \tan(\theta - \alpha))$.

Another source of phase error is that caused by the wet troposphere delay. According to SWOT documentation, the height error caused by the wet troposphere over the ocean is 0.42 cm if the microwave radiometer is present and 1.38 cm if there is no microwave radiometer and the path delay must be estimated algorithmically [4]. Since the wet troposphere delay is given in distance units, it is included directly in the error equation as σ_o. Eq. (6.26) can

be combined with the baseline error and the wet troposphere error to obtain the height error using Eq. (6.17). All errors not specified herein were not considered, including the error due to instrument noise. Future work can augment model fidelity by including additional error terms.

It is presumed that the formation undergoes station-keeping to avoid secular drift between the satellites. It is therefore only necessary to average the swath error performance over a single orbit. At discrete points around the orbit, the height error at each point in the swath is calculated. The impact of any in-track separation of the receivers is not considered. The target points are distributed along the ground in cross-track according to the radar field of view. The error at each point in the swath is calculated for each receiver pair, then the lowest error is taken. The error could be reduced further through data synthesis from multiple receiver pairs. The overall swath error is calculated by averaging the error at each point in the swath, then averaging over time.

6.2.3 Optimization formulation

Mission definition

The optimization sets a variety of radar parameters, including transmit frequency, antenna cross-track beamwidth, antenna boresight angle, radar transmit power, pulse repetition frequency, and pulse width. It also sets the inclination and altitude of the circular reference orbit of the formation. The satellites are evenly distributed around a formation that describes an ellipse in the radial/cross-track directions. The maximum cross-track and radial distances dictate the size of the formation's ellipse, as is described in subsection 6.2.2. The semimajor axis of the ellipse is parallel to the radial direction ($\psi = 90$). Because the spacing in the radial direction also defines the in-track spacing, the radial spacing takes a larger range of values

to promote in-track spacing and avoid collisions. Additionally, cross-track spacing requires a difference in inclination, which causes long-term drift between the satellites. Future analyses could permit larger cross-track separations and calculate the fuel needed for stationkeeping. The optimization determines between one and six satellite configurations, which payloads are included in each configuration, and the number of satellites of each configuration in the formation. A configuration is defined as the set of payloads and payload parameters that describe a satellite. There may be multiple satellites that have the same configuration in a formation. All configurations are assumed to host a receiver that can receive both VV and HH signals. There are five potential additional payloads: a VV-polarized radar transmitter, an HH-polarized radar transmitter, a nadir altimeter, a microwave radiometer, and a downlink communications transmitter.

With the exception of the radar transmitters, radar receivers, and downlink communications units, multiple instances of the payloads provide no benefit beyond redundancy. At least two receivers are required to perform interferometry, and additional receivers improve baseline consistency. Additionally, a team at the German Aerospace Center has suggested that an increased number of receivers flying in close formation could be used to reduce the required pulse repetition frequency of the radar, which reduces the average power used by the system and increases the allowable swath width. This approach has not been considered in this study. The same study suggested that multiple transmitters could alternate pulses, again decreasing the average power [109]. It is instead assumed that any redundant transmitters will alternate illuminating the ground over longer periods of time, which has the same effect on average power. Simultaneous illumination by the transmitters could lead to a lower required peak power but would create complexity in processing and is therefore not considered. The X-band communications unit can be included on multiple satellites to decrease the communications payload power requirement for each satellite.

The radar bandwidth is assumed to be 200 MHz. The two transmitter types are assumed to have orthogonal polarizations (VV and HH). The in-track dimension of the antenna is fixed at 5 m. The radar antennas are assumed to be reflectarray antennas. The interferograms are assumed to use multilook processing with a posting of 250 m prior to transmission to ground.

The nadir altimeter is assumed to be identical to the Poseidon altimeter flown on the Jason-3 mission, which has a mass of 70 kg and a power draw of 78 W. Likewise, the microwave radiometer is assumed to have the mass and power draw of the AMR flown on the Jason missions, 27 kg and 31 W. However, the planned design for SWOT's microwave radiometer includes non-nadir facing channels to better observe the off-nadir water vapor and reduce swath error. This additional level of complexity, which may increase mass and power requirements, is not considered in the current optimization. The precision orbit determination (POD) hardware is not modelled separately in the current analysis. POD hardware, particularly GPS receivers, is standard on monolithic satellites, so it is assumed that it is included in the historic mass, power, and cost estimates used in this analysis.

The crosslink hardware between the satellites is not explicitly modelled. Optical crosslinks can be used to transfer the data with a minimal increase in mass. Synchronization between the satellites is not considered. The planned mission lifetime is five years. The decision variables for the optimization are shown in Table 6.1.

A truly monolithic configuration is not considered in this analysis. However, a monolith could easily be considered in the analysis by reducing the minimum number of satellites from three to one and by including a decision variable defining a rigid baseline for collocated radar receivers. The error functions for the monolith are different because the baseline and roll error are a function of rigidity and attitude knowledge rather than relative position.

Table 6.1: InSAR formation decision variables.

DV	Description	Lower limit	Upper limit	Type

x_1	Days until ground track repeats	21	23	$\mathbb{Z}$
x_2	Number of satellite formations	1	4	$\mathbb{Z}$
x_{i3}	Number of satellites of the i-th type in the formation	1	5	$\mathbb{Z}$
y_1	Antenna cross-track half-power beamwidth (HA_c), °	2	20	$\mathbb{R}$
y_2	Antenna cross-track boresight angle (β_c), °	3	30	$\mathbb{R}$
y_3	Radar peak transmit power (P), W	100	2,000	$\mathbb{R}$
y_4	Altitude H, km	300	2,000	$\mathbb{R}$
y_5	Inclination (i), °	60	120	$\mathbb{R}$
y_6	Radar center frequency (f), GHz	1	40	$\mathbb{R}$
y_7	Formation cross-track distance from center (m) (b_e), m	0.1	70	$\mathbb{R}$
y_8	Formation radial-track distance from center (m) (a_e), m	5	4,000	$\mathbb{R}$
y_9	Pulse repetition frequency (PRF), Hz	500	6,000	$\mathbb{R}$
y_{10}	Pulse width (τ), s	10^{-6}	10^{-4}	$\mathbb{R}$
y_{i11}	VV-transmitter inclusion	0	1	$\mathbb{R}$
y_{i12}	HH-transmitter inclusion	0	1	$\mathbb{R}$
y_{i13}	Downlink communications inclusion	0	1	$\mathbb{R}$
y_{i14}	Microwave radiometer inclusion	0	1	$\mathbb{R}$
y_{i15}	Nadir altimeter inclusion	0	1	$\mathbb{R}$
$i_{\max}$	Number of satellite configurations	1	6	$\mathbb{Z}$
for i	Satellite configuration	1	$i_{\max}$	$\mathbb{Z}$

Optimizer

The following optimizer parameters are used.

- **Optimizer:** Variable-length genetic algorithm with epsilon boxing
- **Initial population:** 400
- **Tournament competitors:** 2
- **Mutation rate:** 5%
- **Length change rate:** 10%
- **Expansion rate:** 0%
- **Nondominated generations for convergence:** 20
- **Number of restarts:** 10
- **Maximum archive size:** 10,000
- **Epsilon boxes:** [0.02, 100, 0.05, 0.02]
- **Objective limits:** [5, 2,000, Inf, Inf]

Dynamics

The following dynamics parameters are used.

- **Propagator:** Fast J4 propagator
- **Central body:** Earth
- **Scenario epoch:** Oct 9, 2019, 00:00
- **Scenario duration:** 23 days
- **Propagation time step:** 86.4 sec
- **Satellite pointing:** Nadir

Constraints

Eleven constraints are used in this scenario for the optimization.

Constraint #1, repeating ground track orbit: As required by the SWOT program, the reference orbit for the formation must be a repeating ground track orbit. Constraint formulation for repeating ground tracks is described in subsection 3.5.4. Adapted from Eq. (3.39), the constraint equation is

$$h_1(\mathbf{x},\, \mathbf{y}) = mT_S(\mathbf{y}) - x_1 T_G(\mathbf{y}) = 0, \qquad m \in \mathbb{Z} \tag{6.27}$$

where T_S is the nodal period of the reference orbit and T_G is the nodal period of a ground point at the equator. This constraint is repairable. If the value for m that satisfies Eq. (6.27) is not integer, it is rounded to the nearest integer. The semimajor axis of the orbit is then modified so the constraint equation is satisfied.

Constraints #2–4, SAR limitations: There are several constraints native to pulsed radar performance that should be satisfied to prevent overlapping echoes [36].

1. The pulse width must be sufficiently short that the radar finishes transmitting before it begins to receive. If r is the distance to the target, c is the speed of light, and τ is the pulse width, $\tau \leq \frac{2r}{c}$. The constraint is

$$g_1(\mathbf{y}) = y_{10} - \frac{2r(\mathbf{y})}{c} \leq 0 \tag{6.28}$$

   This constraint is repairable. If a solution violates this constraint, the pulse width is adjusted.

2. Sampling occurs at a minimum of twice the highest Doppler bandwidth frequency. Where v is the radar speed and d_a is the in-track diameter, $\frac{2v}{d_a} \leq \text{PRF}$. The constraint is

$$g_2(\mathbf{y}) = \frac{2v(\mathbf{y})}{d_a} - y_9 \leq 0 \tag{6.29}$$

   This constraint is repairable. If a solution violates this constraint, the PRF is adjusted.

3. The swath width is small enough that the time between when the near side of the swath is returned and when the far side of the swath is returned is less than the pulse repetition interval. Where H is the platform altitude, θ_1 is the near-edge look angle, and θ_2 is the far-edge look angle, $\text{PRF} \leq c/(2H(\sec(\theta_2) - \sec(\theta_1)))$. The constraint is

$$g_3(\mathbf{y}) = y_9 - \frac{c}{2y_4(\sec(y_2 + y_1/2) - \sec(y_2 - y_1/2))} \leq 0 \tag{6.30}$$

   This constraint is repairable. If a solution violates this constraint, the antenna half-power beamwidth is adjusted.

4. The echo should be received while the radar is not transmitting (but not necessarily the same window when it was transmitted). This constraint is not explicitly modelled. As a result, the PRF may not be valid for the generated design solution and may require tweaking.

These constraints do not consider the increase in path length due to the nonzero in-track position of the satellites relative to the ground targets, which would increase the allowable pulse width. Because the in-track distance would be similar across the entire swath, the difference in cross-track distances has less impact on the overall path length. Therefore, the difference between path lengths at the near and far edge of the swath would decrease, raising the upper limit on the PRF. As the addition of an in-track distance from the targets relaxes the constraints, the case where the ground targets are confined to the cross-track/radial frame of the satellite is more stringent.

Constraints #5–6, number of satellites per formation: There must be at least three satellites in the formation to avoid points in the orbit at which the perpendicular baseline goes to zero. A maximum of ten satellites are allowed per formation. The constraints are

$$g_4(\mathbf{x}) = 3 - \sum_{i=1}^{i_{\max}} x_{i3} \leq 0 \tag{6.31}$$

$$g_5(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i3} - 10 \leq 0 \tag{6.32}$$

The constraint is repairable. The number of satellites of each type are randomly modified until the constraints are satisfied.

Constraints #7–11, payload inclusion: Each of the payloads must be included at least

once. Adapted from Eq. (3.37), the constraints are

$$g_6(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i11} \rceil \leq 0 \tag{6.33}$$

$$g_7(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i12} \rceil \leq 0 \tag{6.34}$$

$$g_8(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i13} \rceil \leq 0 \tag{6.35}$$

$$g_9(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i14} \rceil \leq 0 \tag{6.36}$$

$$g_{10}(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i15} \rceil \leq 0 \tag{6.37}$$

The constraint is repairable. If a payload is not present, it is randomly added to one of the configurations.

Objectives

The optimization has four objectives.

Objective #1, minimize average sea surface height error: The sea surface height measurement error is calculated as described in subsection 6.2.2. The objective function is

$$f_1(\mathbf{x}, \mathbf{y}) = \frac{1}{n_t n_g} \sum_{i=1} n_t \sum_{j=1} n_g \sigma_h(\mathbf{x}, \mathbf{y}) \tag{6.38}$$

where the error function $\sigma_h(\mathbf{x}, \mathbf{y})$ from Eq. (6.17) is averaged over time and points in the radar field of view. n_t is the number of time steps and n_g is the number of target points.

Objective #2, minimize cost: Satellite mass is estimated according to the sizing methodology described in section 3.4. The overall mass and subsystem mass estimates are then used to estimate satellite and launch costs according to the cost methodology described in sub-

section 3.6.2. The radar payload is assumed to have a technology readiness level (TRL) of four for the first flight unit and seven for all subsequent units. For low-mass satellites, DISCO-Tech uses SSCM to estimate most costs and NICM to estimate the radar transmitter cost. All other payload costs are estimated as 40% of the bus cost, as SSCM does not limit payload type. Because USCM is designed for communications payloads, NICM is used to estimate the cost for both the radar transmitters and the nadir altimeter for high-mass satellites. The other payload costs are estimated according to their power usage as described by USCM. The learning rate is estimated as 0.95. Adapted from Eq. (3.77), the objective function is

$$f_2(\mathbf{x},\mathbf{y}) = LC(\mathbf{x},\mathbf{y}) + \sum_{i=1}^{i_{\max}} NRC_i(\mathbf{x},\mathbf{y}) + RC_i(\mathbf{x},\mathbf{y}) * (x_2 x_{i3})^{1+\log_2(0.95)} \tag{6.39}$$

where $LC(\mathbf{x},\mathbf{y})$ is the launch cost, $NRC_i(\mathbf{x},\mathbf{y})$ is the nonrecurring cost for the i-th satellite type, and $RC_i(\mathbf{x},\mathbf{y})$ is the recurring cost for the i-th satellite type.

Objective #3, minimize equatorial revisit time: Since the fields of view for SAR satellites are typically small, it is not practical to calculate coverage using a target grid. Instead, spherical geometry is used to calculate the longitude band covered by the sensors on a satellite when it passes over the equator. The longitudes beneath the satellite are then propagated forward by half of the satellite's nodal period, $T_S = 2\pi/(n + 1.5 J_2 n r_E^2 a^{-2}(3 - 4\sin^2 i))$ to the next equatorial crossing. n is the mean motion, r_E is the radius of the planet, and J_2 is the Earth's J_2 coefficient. The longitudes that have not yet been overflown are tracked. If multiple planes/formations of satellites are used, it is assumed that the formations are arranged in such a way as to maximize coverage by minimizing overlapping ground traces.

When some total percentage of the equatorial longitudes have been overflown, the propagation ceases. For this simulation, the propagation runs until 99% coverage at the equator has

been achieved or until 23 days have passed. The objective value is the negative of the percentage of the allotted days remaining after the desired level of coverage has been achieved. If the desired coverage is not achieved, the objective value is the percentage of area covered. The objective function is

$$f_3(\mathbf{x},\mathbf{y}) = \delta_c(\mathbf{x},\mathbf{y})(T_r(\mathbf{x},\mathbf{y}) - 23)/23 + (1 - \delta_c(\mathbf{x},\mathbf{y}))A_c(\mathbf{x},\mathbf{y}) \tag{6.40}$$

where δ_c is one if sufficient coverage is achieved in the allowed time period and zero otherwise. $T_r(\mathbf{x},\mathbf{y})$ is the time needed in days to achieve sufficient coverage. $A_c(\mathbf{x},\mathbf{y})$ is the difference between 99% and the fraction of the equator covered in 23 days. The number of days required for the satellites to cover 99% of the equator will hereafter be referred to as *time to 99% coverage* or *time to coverage* to distinguish this metric from the equatorial maximum revisit time, which would require coverage of 100% of equatorial longitudes.

Objective #4, maximize resilience to expected failure: Because the satellites are flying in formation rather than providing distributed coverage, a coverage-based resilience analysis is not appropriate in this scenario. Instead, the impact and likelihood of a satellite loss on the ability to measure SSH is assessed. Because the nadir altimeter provides redundancy and validation and is not strictly necessary for InSAR, its impact is not considered. Additionally, while the microwave radiometer improves the accuracy of the measurements, mathematical modelling can be used to estimate the path delays in the atmosphere. The difference in accuracy is about 1.38 cm, but the mission is not significantly impacted. Therefore, the impact of the loss of the radiometer is also not considered.

The remaining subsystems are the radar transmitters, radar receivers, and downlink communications. If all of the transmitters of one type are lost, half of the covered area is lost and the performance reduced by 50%. The increased load on identical transmitters due

to one or more losses is not considered as ground targets can be alternated to reduce the operational load. If there are fewer than two satellites, performance is reduced to 0% as no simultaneous interferometric images can be formed. If a downlink communications unit is lost, performance degrades as the full data load cannot be transmitted to the ground.

The performance is calculated for all combinations of satellite losses in the formation, ignoring the case where no satellite is lost. This formulation is an extension of the one used in Ref. 96, which considers only a single satellite loss and does not consider failure rates. The probability of failure for each satellite is calculated according to the Weibull distribution described in subsection 3.6.2. The expected performance is calculated and multiplied by -1 to get the objective value. The objective function is

$$f_4(\mathbf{x},\mathbf{y}) = -\frac{\sum_{k=1}^{n_c} P_k(\mathbf{x},\mathbf{y}) p_k(\mathbf{x},\mathbf{y})}{\sum_{k=1}^{n_c} p_k(\mathbf{x},\mathbf{y})} \tag{6.41}$$

where n_c is the total number of failure combinations, P_k is the relative performance of the k-th degraded formation, and p_k is the probability that the k-th formulation will occur. Performance is assessed on a formation level, so the number of formations of satellites does not affect the resilience analysis.

Mathematical summary

The mathematical formulation for the optimization is shown below.

Minimize

$$f_1(\mathbf{x},\ \mathbf{y}) = \sigma_h(\mathbf{x},\ \mathbf{y})$$

$$f_2(\mathbf{x},\mathbf{y}) = LC(\mathbf{x},\mathbf{y}) + \sum_{i=1}^{i_{\max}} NRC_i(\mathbf{x},\mathbf{y}) + RC_i(\mathbf{x},\mathbf{y}) * (x_2 x_{i3})^{1+\log_2(0.95)})$$

$$f_3(\mathbf{x},\mathbf{y}) = \delta_c(\mathbf{x},\mathbf{y})(T_r(\mathbf{x},\mathbf{y}) - 23)/23 + (1 - \delta_c(\mathbf{x},\mathbf{y}))A_c(\mathbf{x},\mathbf{y})$$
$$f_4(\mathbf{x},\mathbf{y}) = -\frac{\sum_{k=1}^{n_c} P_k(\mathbf{x},\mathbf{y})p_k(\mathbf{x},\mathbf{y})}{\sum_{k=1}^{n_c} p_k(\mathbf{x},\mathbf{y})}$$

such that

$$i_{\max} \in [1,6] \cap \mathbb{Z} \qquad y_{i13} \in [0,1]\ \forall i = 1,2,\ldots,i_{\max}$$
$$x_1 \in [21,23] \cap \mathbb{Z} \qquad y_{i14} \in [0,1]\ \forall i = 1,2,\ldots,i_{\max}$$
$$x_2 \in [1,4] \cap \mathbb{Z} \qquad y_{i15} \in [0,1]\ \forall i = 1,2,\ldots,i_{\max}$$
$$x_{i3} \in [1,5] \cap \mathbb{Z}\ \forall i = 1,2,\ldots,i_{\max} \qquad g_1(\mathbf{y}) = y_{10} - \frac{2r(\mathbf{y})}{c} \leq 0$$
$$y_1 \in [2,20] \qquad g_2(\mathbf{y}) = \frac{2v(\mathbf{y})}{d_a} - y_9 \leq 0$$
$$y_2 \in [3,30] \qquad g_3(\mathbf{y}) = y_9 - \frac{c}{2y_4} *$$
$$y_3 \in [100,2,000] \qquad (\sec(y_2 + y_1/2) - \sec(y_2 - y_1/2))^{-1} \leq 0$$
$$y_4 \in [300,2,000] \qquad g_4(\mathbf{x}) = 3 - \sum_{i=1}^{i_{\max}} x_{i3} \leq 0$$
$$y_5 \in [60,120] \qquad g_5(\mathbf{x}) = \sum_{i=1}^{i_{\max}} x_{i3} - 10 \leq 0$$
$$y_6 \in [1,40] \qquad g_6(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i11} \rceil \leq 0$$
$$y_7 \in [0.1,70] \qquad g_7(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i12} \rceil \leq 0$$
$$y_8 \in [5,4,000] \qquad g_8(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i13} \rceil \leq 0$$
$$y_9 \in [500,6,000] \qquad g_9(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i14} \rceil \leq 0$$
$$y_{10} \in [10^{-6}, 10^{-4}] \qquad g_{10}(\mathbf{y}) = 1 - \sum_{i=1}^{i_{\max}} \lfloor y_{i15} \rceil \leq 0$$
$$y_{i11} \in [0,1]\ \forall i = 1,2,\ldots,i_{\max} \qquad h_1(\mathbf{x},\ \mathbf{y}) = mT_S(\mathbf{y}) - x_1 T_G(\mathbf{y}) = 0$$

$$y_{i12} \in [0,1] \ \forall i = 1,2,\ldots,i_{\max}$$

6.2.4 Results

The simulation ran for 4,367 generations and produced 77 epsilon-nondominated solutions. The cost and height error objective values for the nondominated front are shown in Fig. 6.6, colored by frequency. From this figure, it is clear that height error increases with increasing frequency. This relationship is a consequence of the fact that the critical baseline is smaller for higher frequencies, as shown in Eq. 6.20. Because the relative position error is assumed to be 1 mm and the roll angle α is dependent on the ratio between the position error and the baseline, a smaller baseline increases the error due to error in the roll angle, raising the total error. Another trend that can be observed is a decrease in error with increasing cost. Increased cost corresponds to increased SAR transmission power. Increased power can improve performance either by decreasing the thermal decorrelation for a fixed distance to the target or by increasing the allowable distance to the target without increasing thermal decorrelation, thereby increasing the critical baseline.

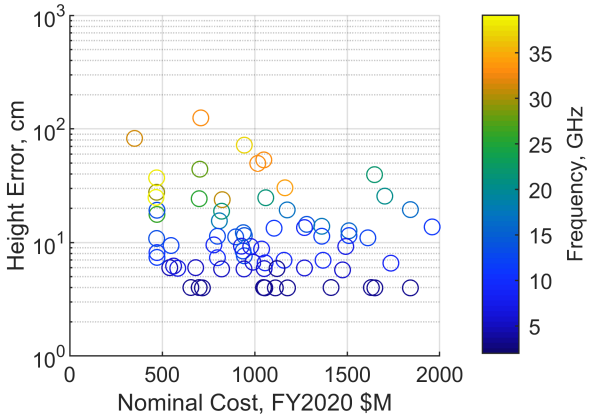


Figure 6.6: Full nondominated front for cost versus height error.

The nondominated front in Fig. 6.6 is sorted into fronts according to the values of the two

objective functions not included in the figure. Fig. 6.7 filters the solution by the time to coverage. Nondominated solutions are found from the set of solutions below various coverage thresholds. The least limiting case considered is the case in which coverage time is less than or equal to 12 days. The various fronts can be used to quantify the increase in cost and error between different coverage thresholds. To quantify these differences for a particular front, each point in the front is considered separately. The closest point on the least limiting front is found that has superior performance in the objective under consideration. The distance in the objective under consideration is found. The average is taken across all points on the current front to find the average shift in the objective, as shown in Table 6.2. These shifts can approximate a "shadow price" for the corresponding coverage threshold. For example, changing the coverage requirement from 12 days to 5 days has a cost of $114M when optimizing over cost and error.

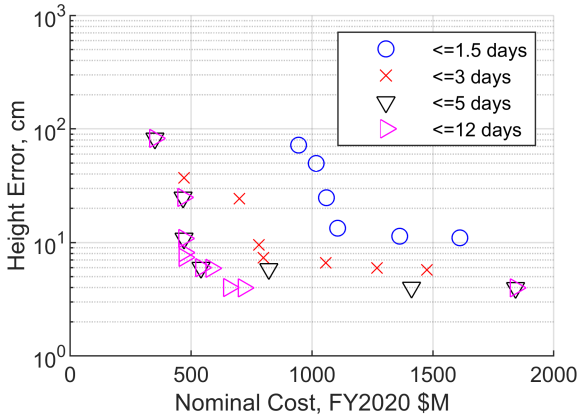


Figure 6.7: Cost versus height error front, filtered by time to coverage.

Table 6.2: Increase in cost and error due to more restrictive time to coverage requirements.

Shift	**Cost increase ($M)**	**Height error increase (cm)**
$\leq$12 days $\rightarrow \leq$ 5 days	114	0.26
$\leq$12 days $\rightarrow \leq$ 3 days	309	5.53
$\leq$12 days $\rightarrow \leq$ 1.5 days	611	18.21

Fig. 6.8 filters the solution by resilience. Nondominated solutions are found from the set of

solutions above various resilience thresholds. The least limiting case considered is the case in which resilience is at least 0.4. The various fronts can be used to quantify the increase in cost and error between different resilience thresholds, as shown in Table 6.3. These shifts can approximate a "shadow price" for the corresponding resilience threshold. For example, a resilience requirement of 0.9 increases cost by $388M compared to a resilience requirement of 0.8 when optimizing over cost and error.

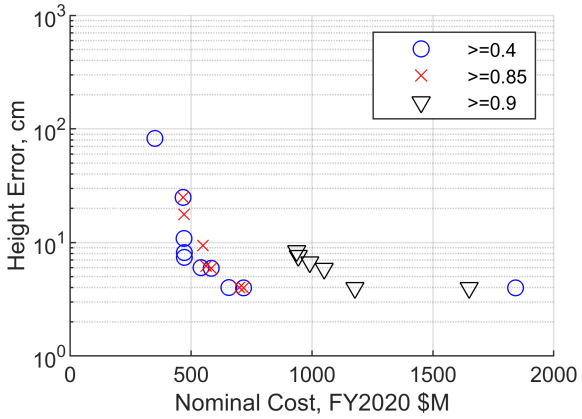


Figure 6.8: Cost versus height error front, filtered by resilience.

Table 6.3: Increase in cost and error due to more restrictive resilience requirements.

Shift	**Cost increase ($M)**	**Height error increase (cm)**
$\geq 0.4 \rightarrow \geq 0.8$	20	1.18
$\geq 0.4 \rightarrow \geq 0.9$	408	2.11

The cost versus time to achieve global coverage is shown in Fig. 6.9 for all points on the nondominated front. There is a trend toward decreasing time to coverage with increased cost. Increased cost can increase the number of satellite formations, increasing the covered area. Alternately, increased cost can increase the transmit power and therefore allow an increase in the satellite field of view for the same Effective Isotropic Radiated Power (EIRP). There is a slight trend towards decreased time to coverage with increased frequency. Because the in-track antenna dimension is fixed in this analysis, the gain due to the in-track field of view

increases with increasing frequency, allowing the cross-track field of view to increase without affecting the total gain.

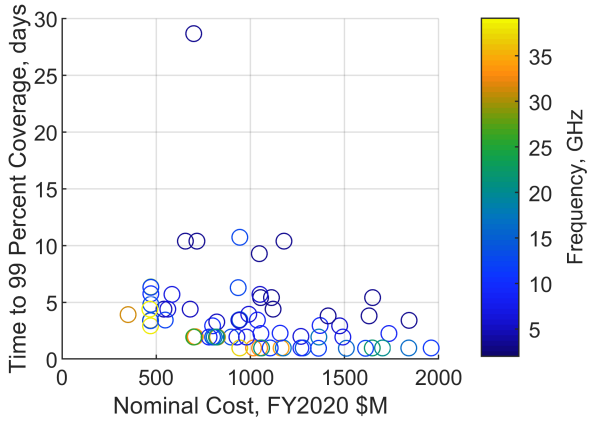


Figure 6.9: Full nondominated front for cost versus time to coverage.

The nondominated front in Fig. 6.9 is sorted into fronts according to the values of the two objective functions not included in the figure. Fig. 6.10 filters the solution by the height error. Nondominated solutions are found from the set of solutions below various error thresholds. The least limiting case considered is the case in which height error is less than or equal to 50 cm. The various fronts can be used to quantify the increase in cost and coverage time between different error thresholds, as shown in Table 6.4. These shifts can approximate a "shadow price" for the corresponding height error threshold. For example, the 10 cm error requirement has a cost of \$71M compared to a 50 cm error requirement when optimizing over cost and coverage. The decrease in cost difference from 10 cm to 7 cm is due to the normalization used to determine the closest point, which treats one day equal to \$100 million.

Fig. 6.11 filters the solution by resilience. Nondominated solutions are found from the set of solutions above various resilience thresholds. The least limiting case considered is the case in which resilience is at least 0.4. The various fronts can be used to quantify the increase in cost and coverage time between different resilience thresholds, as shown in Table 6.5. These shifts

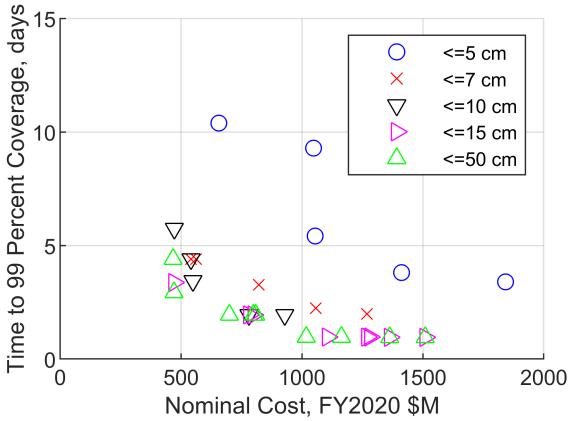


Figure 6.10: Cost versus time to coverage front, filtered by height error.

Table 6.4: Increase in cost and coverage time due to more restrictive height error requirements.

Shift	Cost increase ($M)	Coverage time increase (days)
$\leq$50 cm $\rightarrow \leq$ 15 cm	49	0.06
$\leq$50 cm $\rightarrow \leq$ 10 cm	71	0.37
$\leq$50 cm $\rightarrow \leq$ 7 cm	65	0.72
$\leq$50 cm $\rightarrow \leq$ 5 cm	170	4.61

can approximate a "shadow price" for the corresponding resilience threshold. For example, a resilience requirement of 0.9 increases cost by $128M compared to a resilience requirement of 0.8 when optimizing over cost and coverage.

Table 6.5: Increase in cost and coverage time due to more restrictive resilience requirements.

Shift	Cost increase ($M)	Coverage time increase (days)
$\geq 0.4 \rightarrow \geq 0.8$	26	0.10
$\geq 0.4 \rightarrow \geq 0.9$	154	1.56

The cost versus resilience is shown in Fig. 6.12, colored by the maximum satellite dry mass. Increasing cost has a slight tendency to increase resilience due to the fact that more redundant subsystems or more satellites can be included. There is also a positive correlation between the maximum satellite mass and the resilience. Resilience shows an increase when the mass is above 500 kg. This result is a consequence of the Weibull distributions used to

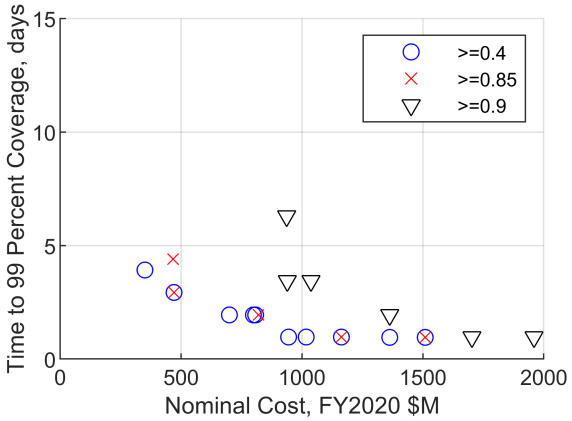


Figure 6.11: Cost versus time to coverage front, filtered by resilience.

model satellite failure rates. For the five year mission lifetime used in this analysis, a satellite weighing less than 500 kg has a failure rate of about 4%. A satellite weighing between 500 and 2,500 kg has a failure rate of about 2%. The satellites with radar transmitters are generally the heaviest in the formation. Since resilience is a function of the relative failure rates of the satellites, transmitting satellites in the 500-2,500 kg mass class are less likely to fail than other satellites in the formation, increasing the resilience.

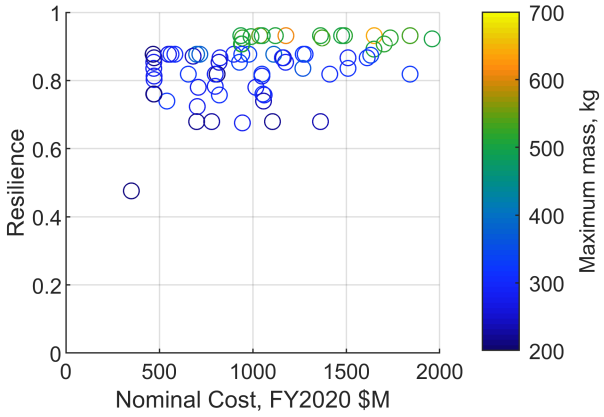


Figure 6.12: Full nondominated front for cost versus resilience.

The nondominated front in Fig. 6.12 is sorted into fronts according to the values of the two objective functions not included in the figure. Fig. 6.13 filters the solution by the

height error. Nondominated solutions are found from the set of solutions below various error thresholds. The least limiting case considered is the case in which height error is less than or equal to 50 cm. The various fronts can be used to quantify the increase in cost and decrease in resilience between different error thresholds, as shown in Table 6.6. These shifts can approximate a "shadow price" for the corresponding height error threshold. For example, the 10 cm error requirement has a cost of $23M compared to a 50 cm error requirement when optimizing over cost and resilience. The decrease in resilience difference from 10 cm to 7 cm is due to the normalization used to determine the closest point, which treats a resilience value of 0.02 equal to $100M.

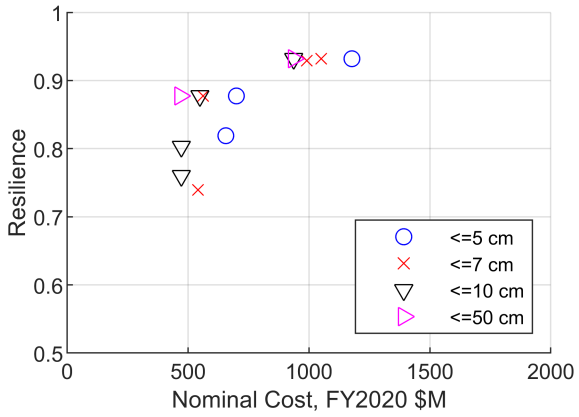


Figure 6.13: Cost versus resilience front, filtered by height error.

Table 6.6: Increase in cost and decrease in resilience due to more restrictive height error requirements.

Shift	**Cost increase ($M)**	**Resilience decrease**
$\leq$50 cm $\rightarrow \leq$ 10 cm	23	0.05
$\leq$50 cm $\rightarrow \leq$ 7 cm	85	0.04
$\leq$50 cm $\rightarrow \leq$ 5 cm	221	0.04

Fig. 6.14 filters the solution by time to coverage. Nondominated solutions are found from the set of solutions above various time to coverage thresholds. The least limiting case considered is the case in which time to coverage is less than or equal to 12 days. The various fronts can

be used to quantify the increase in cost and resilience between different coverage thresholds, as shown in Table 6.7. These shifts can approximate a "shadow price" for the corresponding coverage threshold. For example, changing the coverage requirement from 12 days to 5 days has a cost of $32M when optimizing over cost and resilience.

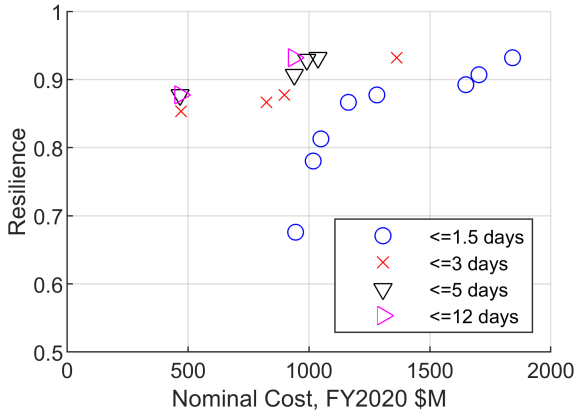


Figure 6.14: Cost versus resilience front, filtered by time to coverage.

Table 6.7: Increase in cost and decrease in resilience due to more restrictive time to coverage requirements.

Shift	**Cost increase ($M)**	**Resilience decrease**
$\leq$12 days $\rightarrow \leq$ 5 days	32	0.01
$\leq$12 days $\rightarrow \leq$ 3 days	305	0.04
$\leq$12 days $\rightarrow \leq$ 1.5 days	512	0.08

The relationship between the level of disaggregation and the resilience of the system is examined. This analysis used resilience to expected failure as an objective, but the resilience to threat of the solutions is considered as well. In this discussion, "monolithic" refers to a system in which all five assignable payloads are collocated on a single satellite. Such solutions also have small, receive-only satellites and are therefore not truly monolithic. These solutions are the most aggregated systems possible for the optimization due to the requirement that at least three satellites be present. Disaggregated systems without redundancy have lower resilience to expected failure than monolithic systems because there are several

satellites whose failure would degrade performance. In the current modelling methodology, satellite failure rates within a given mass class are dependent only on mission lifetime. Therefore, the increase in mission-critical satellites reduces resilience to expected failure. For example, a formation with one downlink satellite, one VV-transmitting satellite, and one HH-transmitting satellite would have no performance if either the downlink failed or both the VV and HH satellites failed. The formation would have half performance if either the VV or the HH satellite failed. If the satellites have a failure rate of 4%, the probabilities that the formation will have no, half, and full performance are 4.16%, 7.37%, and 88.47% respectively. In this case, the formation will have some capability 95.84% of the time. The vulnerability due to the single downlink unit highlights a weakness in disaggregated systems that rely on communications through a single satellite. However, the simplicity of the satellites in the disaggregated architecture reduces the cost and time to construction if a module must be replaced.

Collocating downlink systems with radar transmitters raises resilience to expected failure slightly. The probabilities that the formation will have no, half, and full performance are 0.16%, 7.68%, and 92.16% respectively. In this case, the formation will have some capability 99.84% of the time. Although the likelihood of full performance is less than that of the monolithic case discussed below, the likelihood that there will be at least some level of performance is higher than in the monolithic case. This result indicates that graceful degradation is more likely with disaggregated systems.

Disaggregated systems with redundancy in the radar transmitters have increased resilience to failure but are significantly more costly. Because the downlink units are sized according to the number of downlink units in the system, increasing the number of downlink units does not increase resilience significantly. Using redundant downlink units that can individually handle the entire formation's data burden would increase resilience.

Conversely, combining the downlink, VV transmitter, and HH transmitter onto a single satellite has a higher chance of complete failure but also a higher chance of no failure. The probability that this monolithic satellite will have no performance is 4%, and the probability that the satellite will have full performance is 96%. Additionally, more massive satellites have a lower probability of failure according to the Weibull distributions used to calculate failure rate [46]. However, these mass-based failure rates are based on historic performance. Reduced testing, less reliable components, components without radiation hardening, level of technology maturity, material selection, designer experience, and reduced design timelines could all contribute to high failure rates of small satellites. A possible area of future study could be to identify cost relationships for component reliability, extent of testing, and level of design effort. These factors could then be used as decision variables that drive cost and satellite failure rate. Also, detailed component level failure rates could change the resilience to failure analysis, since simpler payloads would likely have lower failure rates.

Although resilience to threat was not an optimization objective, it can be examined to quantify benefits of the disaggregated systems. If the resilience to threat for a functionally disaggregated system is defined as the worst-case performance with the removal of a single spacecraft, monolithic systems have no resilience to threat. Disaggregated systems with a single transmitter or with collocated radar transmitters also have no resilience to threat, as the loss of the satellite containing those payloads would remove all capability. Disaggregated systems with redundant radar transmitters or with radar transmitters that are not collocated show higher resilience to threat if there are also redundant communications units. Therefore, the most cost-effective means of increasing resilience to threat is to place the radar transmitters on separate satellites from one another with at least two downlink units in the system. An example configuration would be to have a VV/downlink satellite, an HH/downlink satellite, and a nadir/microwave satellite.

Next, the relationship between the level of disaggregation and the overall mission cost was examined. As noted in section 4.2, the choice of cost model affects the level of disaggregation favored by the optimization. This analysis used a hybrid cost model, where SSCM was used for satellites with a dry mass of less than 400 kg and USCM was used for satellites with a dry mass of greater than 400 kg. NICM was used to estimate the radar payload cost. SSCM was used for most disaggregated solutions because their individual satellite masses were less than 400 kg. The monolithic solutions had higher mass and therefore used USCM.

The hybrid cost model produced solutions that had various disaggregation levels at similar costs. The ten cheapest solutions with height errors of less than 20 cm are shown in Table 6.8. The table shows the indices of the satellites on which each payload is present. The first five solutions all have similar costs with various disaggregation levels ranging from completely disaggregated (one payload per satellite) to a solution where the VV transmitter, HH transmitter, and downlink unit are collocated and the nadir altimeter and microwave radiometer are collocated. The number of receive-only satellites varies, but the low cost of the receive-only satellites is unlikely to significantly affect the ordering. The least costly monolithic system has a predicted cost of $936M, 99% higher than the lowest-cost disaggregated solution. This higher cost is due in part to the use of the USCM cost model. Key cost drivers are the number of formations used and the radar transmit power.

If SSCM is used to evaluate the cost for all satellites, the predicted cost of the monolithic system decreases to $626M, making it the tenth-least costly solution with a height error of less than 20 cm. If USCM is used instead to evaluate all satellite costs, the monolithic solution becomes the cheapest option at $1B. All predicted costs are increased. The fully disaggregated system cost increases to $1.2B, likely due to higher development costs. At $1.2B, the fully disaggregated system is the tenth least-costly solution with a height error of less than 20 cm. The ten cheapest USCM-predicted solutions with height errors of less

Table 6.8: Payload distribution for the cheapest systems, hybrid cost model.

Rank	VV	HH	Dwn	Mic	Ndr	Rx	Cost ($M)
1	5	4	3	2	1	6–10	470
2	3	3	3	2	1	4–9	471
3	2	2	2	1	1	3–6	471
4	4	3	2	2	1	5–10	472
5	3	2	2–3	1	1	4–9	472
6	6	5	2–4	1	1	7–9	541
7	3	3	3	2	1	4–10	548
8	3	3	3	2	1	4–10	562
9	2	2	2	1	1	3–10	583
10	4	3	3–4	2	1	5–10	656

than 20 cm are shown in Table 6.9. The relatively close costs and discrepancies between the cost models indicate that various levels of disaggregation should be assessed using detailed, component-level cost modelling before a decision is made as to which level of disaggregation is most appropriate for a mission. It can be seen that the fifth-least costly solution is the same for both cost models, a system with a VV/downlink satellite, an HH/downlink satellite, a nadir altimeter/microwave radiometer satellite, and six receive-only satellites. The consistently low cost for this solution makes it a good candidate for further exploration.

Table 6.9: Payload distribution for the cheapest systems, USCM.

Rank	VV	HH	Dwn	Mic	Ndr	Rx	Cost ($M)
1	1	1	1	1	1	2–7	1000
2	1	1	1	1	1	2–10	1002
3	1	1	1	1	1	2–7	1007
4	1	1	1	1	1	2–9	1021
5	3	2	2–3	1	1	4–9	1025
6	2	2	2	1	1	3–6	1055
7	1	1	1	1	1	2–10	1114
8	3	3	3	2	1	4–9	1126
9	4	3	2	2	1	5–10	1167
10	5	4	3	2	1	6–10	1201

Trade-offs

The ability to quantify trade-offs between objectives is critical when selecting a design from the nondominated front or when attempting to identify patterns in the objective space for a mission. For example, the trade-offs between cost and other objectives can inform designers as to where money might be best spent. General trends for trade-offs between cost and other objectives are discussed above for the set of epsilon-nondominated solutions found by the optimizer. Additionally, some numeric information can be gleaned by examining the nondominated front in two objectives when another objective is used as a filter. However, the filtering process can consider only the impact of one objective on two others. If functions can be developed relating the objectives, the gradients of the functions would quantify the rates of change between objectives for an arbitrary number of objectives. To this end, it is desirable to find a hypersurface that approximates the points on the nondominated front. If the data is discontinuous or if it is difficult to find an appropriate hypersurface fit, a numerical approximation of the gradient may be used instead. Two hypersurfaces will be examined: the surface in 3D space relating cost and height error to coverage and the surface in 4D space relating height error, coverage, and resilience to cost.

To increase the number of data points available for the hypersurface fit, the epsilon-dominance criterion was discarded. Function evaluations made across 832 generations were combined with the epsilon-nondominated front to obtain a nondominated front with 3,019 solutions. Points for which the time to achieve coverage was greater than 15 days were removed due to a scarcity of data in this region of the objective space. Additionally, points with measurement errors greater than 1 m or cost greater than $3.5B were considered outliers and removed. The resulting nondominated front has 2,623 solutions.

Coverage is independent of resilience, as resilience in this analysis is only a function of

redundant sensors and sensor distribution. Sensor distribution within a formation does not affect coverage because all formations are required to have radar transmitters that cover a swath on each side of the ground track. Therefore, when analyzing the impact of cost and measurement error on coverage, resilience was ignored. By removing resilience as an objective, a nondominated front with 623 points was generated. This analysis examines the trade-offs between two performance objectives (height error and time to coverage) as well as evaluating the impact of cost on time to coverage.

A fit of the form $z = c_1 + c_2 x^{c_3} + c_4 y^{c_5} + c_6 x^{c_7} y^{c_8}$ was found using nonlinear least squares regression, where x is measurement error, y is cost, and z is time to coverage. The fit is not valid for high cost, high error solutions, since there are no solutions in that region of the objective space on the front. The fit is shown with the data points in Fig. 6.15. The average error of the fit is 1.1 days, and the median error of the fit is 0.7 days. The fit tends to overestimate the time to coverage for low-cost solutions. A contour plot of the estimated time to coverage is shown in Fig. 6.16.

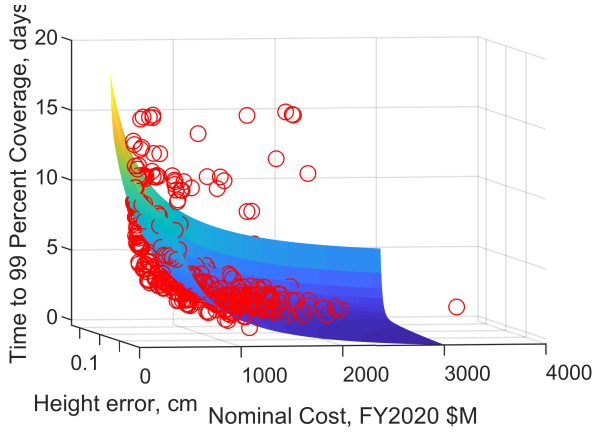


Figure 6.15: Surface fit for time to coverage as a function of cost and error.

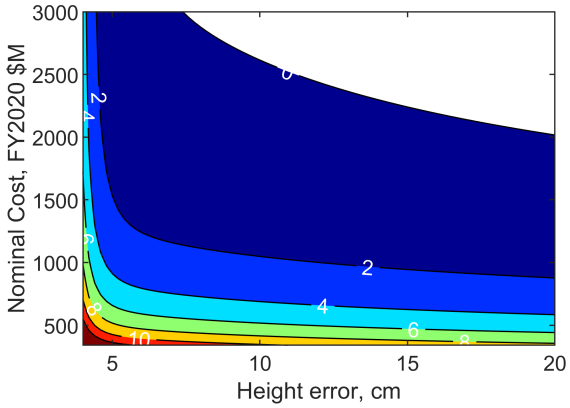


Figure 6.16: Contour plot of estimated time to coverage as a function of cost and error.

The rate of change of time to coverage with respect to cost is shown in Fig. 6.17 with units of days per million dollars. The plot indicates diminishing returns with increasing investment:

increasing spending for a solution with an initial cost of $350M by $100M is expected to improve the time to coverage by 3 days, while the same increase for a solution with an initial cost of $2B would only improve time to coverage by 0.1 days. This result is due to a need to add satellites or expand the radar field of view (and therefore increase radar transmit power) to continually improve time to coverage. Additionally, the return on investment is slightly larger for solutions with low error. For a fixed number of satellites and constant cost, solutions with lower error require narrower fields of view to increase gain and therefore SNR. These solutions therefore have higher times to coverage, so there is more potential for improvement. For example, Fig. 6.16 shows that doubling the solution cost from $500M to $1B at an error of 4 cm decreases the time to coverage from 13 to 8 days, while the same change at 20 cm decreases the time to coverage from 5 to 2 days.

The rate of change of time to coverage with respect to height error is shown in Fig. 6.18 with units of days per centimeter. Increased error can be traded for improved coverage. At 5 cm of initial error and an initial cost of $1B, the rate of change is about -1.3 days/cm, so an error increase of 0.1 cm is expected to improve time to coverage by about 0.13 days. Improvements diminish with increasing cost and increasing error. For the same initial cost with an initial error of 12 cm, the rate of change is -0.08 days/cm.

The set of 2,623 nondominated solutions was used to generate a hypersurface fit relating measurement error, time to coverage, and resilience to cost. A fit of the form $q = c_1 + c_2 x^{c_3} + c_4 y^{c_5} + c_6 z^{c_7} + c_8 (xy)^{c_9} + c_{10} (xz)^{c_{11}}$ was found using nonlinear least squares regression, where x is measurement error, y is time to coverage, z is resilience, and q is cost. The fit does not apply for high error, high time to coverage solutions due to the lack of data points in this region of the objective space. Two slices of the hypersurface are shown in Figs. 6.19 and 6.20 for resilience levels of 0.5 and 0.9, respectively. The fit has a mean percent error of 14.7% and a median percent error of 11.6%. The mean absolute error is $151M and the median

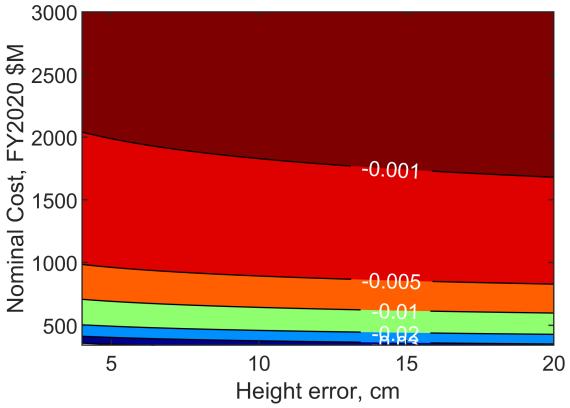


Figure 6.17: Contour plot of rate of change of time to coverage with respect to cost.

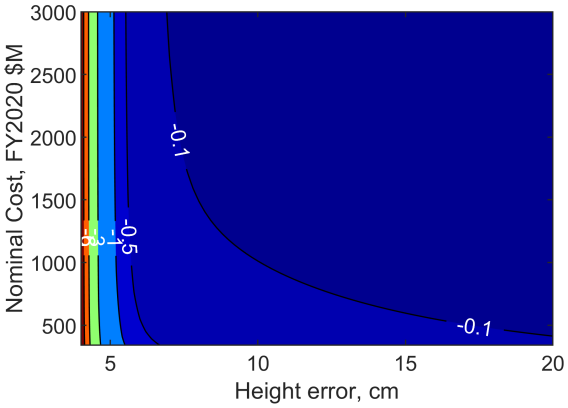


Figure 6.18: Contour plot of rate of change of time to coverage with respect to error.

absolute error is \$108M. The impact of each objective on cost is evaluated by calculating the gradient of the hypersurface across a set of evenly space points.

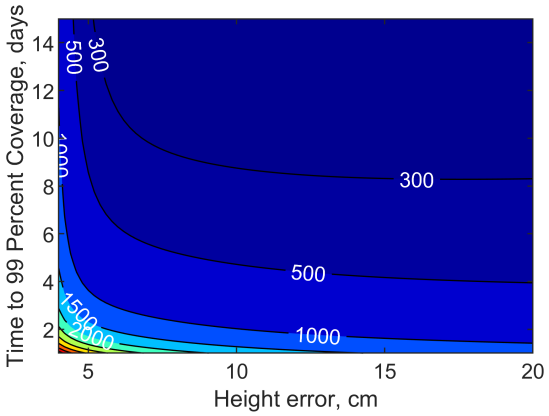


Figure 6.19: Contour plot of estimated cost as a function of error and time to coverage, fixed resilience of 0.5.

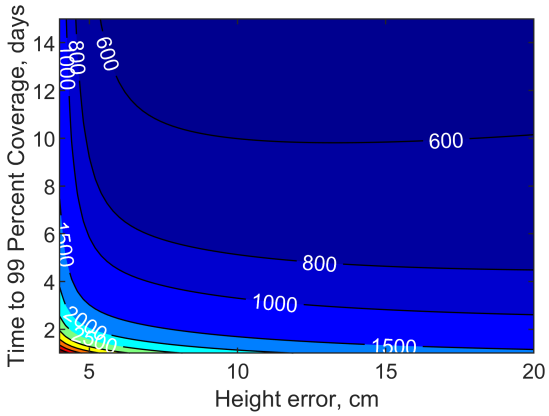


Figure 6.20: Contour plot of estimated cost as a function of error and time to coverage, fixed resilience of 0.9.

Fig. 6.21 shows the rate of change of cost with respect to resilience in units of millions of dollars. The rate of change increases with increasing resilience, meaning that improvements in resilience become more costly as resilience increases. The cost to increase resilience is nearly constant with respect to height error. Because resilience to failure is calculated for all combinations of satellite failures discounting the case where no satellites fail, adding noncrit-

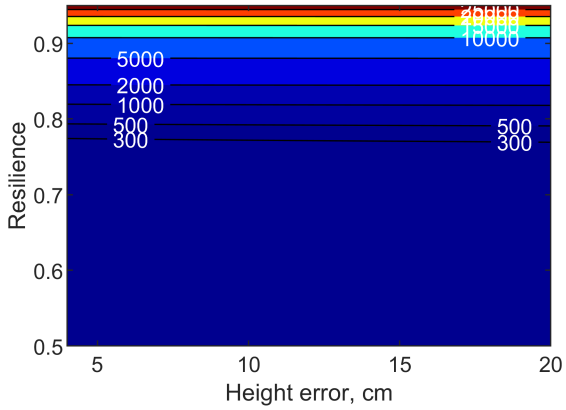


Figure 6.21: Contour plot of rate of change of cost with respect to resilience.

ical satellites increases the resilience to failure. An increase in noncritical satellites increases the number of combinations of satellites for which failures do not affect performance. *Therefore, the cheapest way to add resilience is to increase the number of receive-only satellites.* Even higher resilience may be achieved by aggregating payloads, which further increases cost. Since the cost of the receive-only satellites and the level of aggregation are largely independent of other design parameters, the increase in cost is independent of the height error.

Fig. 6.22 shows the rate of change of cost with respect to time to achieve coverage in units of $M/day. As was discussed in the analysis of the coverage rates of change, decreasing time to coverage increases costs, with the magnitude of the rate of change increasing as the time to coverage decreases.

Fig. 6.23 shows the rate of change of cost with respect to height error in units of $M/cm. Resilience is fixed at 0.9. The contour plot shows that costs increase with decreasing height error. Because the fit is not valid for high error, high time to coverage solutions, the values produced in the upper right corner of the plot are ignored. Costs increase more sharply as the height error decreases. The cost appears to approach an asymptote near 4 cm. This

threshold is a result of the errors due to GPS error and wet tropospheric path delay error, which are fixed based on the assumptions of the analysis.

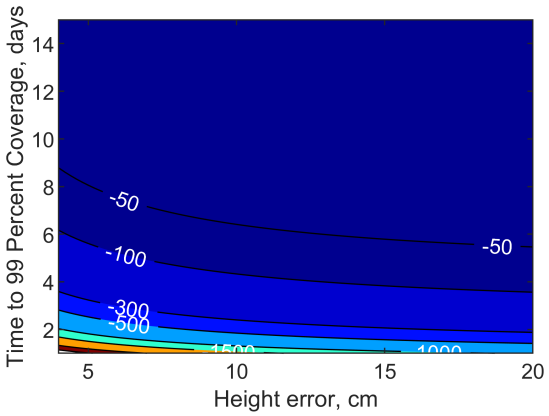


Figure 6.22: Contour plot of rate of change of cost with respect to time to coverage.

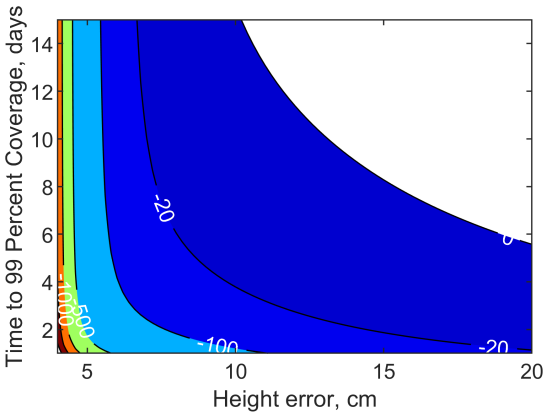


Figure 6.23: Contour plot of rate of change of cost with respect to height error.

The contour plots in Figs. 6.21–6.23 can be used to estimate performance changes based on budget changes. As an example, a reference solution with a height error of 10 cm, a time to coverage of 6 days, and a resilience of 0.9 is considered. From the hypersurface, the expected cost of this solution is \$744M. If the budget for the mission is increased by \$100M, the expected change in performance would be:

- A decrease in height error of 4 cm **OR**
- A decrease in time to coverage of 1.5 days **OR**
- An increase in resilience of 0.011.

If the budget was instead decreased by \$20M, the expected change in performance would be:

- An increase in height error of 4 cm **OR**
- An increase in time to coverage of 0.4 days **OR**

- A trivial decrease in resilience

This analysis can therefore assist in budgetary decision making. Due to the estimated nature of the hypersurface, a solution might not exist at a particular point in the objective space. Either the closest point on the nondominated front can be found or an additional optimization performed with focus on the region of interest in the objective space.

Downselection

To select a single solution for implementation, the MCDM process described in Ref. 64 is used. The MCDM process is applied after removing undesirable solutions. First, all solutions with cost greater that $2 billion are eliminated. Next, all solutions that do not achieve 99% coverage within 23 days are removed. Solutions with error greater than 10 cm are removed due to SWOT's height error requirements. Solutions with inclinations less than 75° or greater than 105° are removed to improve coverage in arctic regions. Finally, solutions with frequencies below 8 GHz are removed because the Shuttle Radar Topography mission saw better returns from the ocean with X-band than with C-band [131]. After this downselection process, 10 solutions remain of the original 77. The objective values for these remaining solutions are shown in Figs. 6.24, 6.25, and 6.26.

The selection process described in Ref. 64 is applied to the downselected set of solutions. The solution selected through the MCDM process is marked with a text box listing the objective values in Figs. 6.24, 6.25, and 6.26. The objective values for the Pareto-compromise vector (which does not correspond to an actual solution found by the optimizer)are marked with a black square. The satellite subsystem breakdown is shown in Table 6.10 for the selected solution. The columns indicate the satellite configuration number (Sat Type), the number of satellites of that type in a formation (Num), and the presence of the vertical

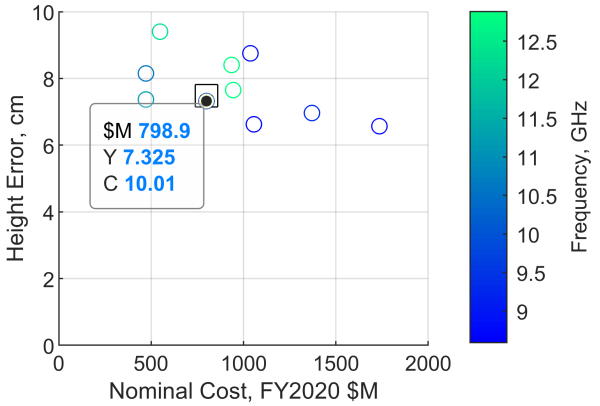


Figure 6.24: Downselected nondominated front for cost versus height error.

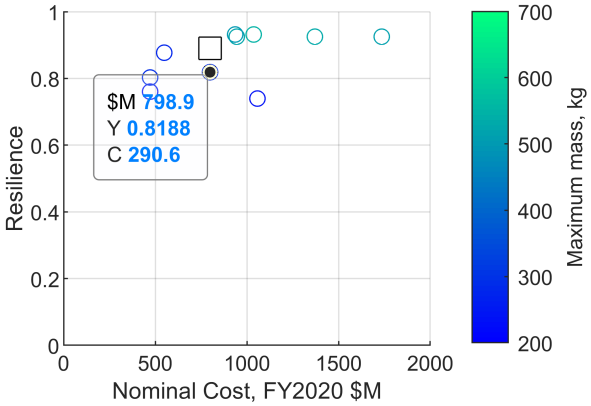


Figure 6.25: Downselected nondominated front for cost versus resilience.

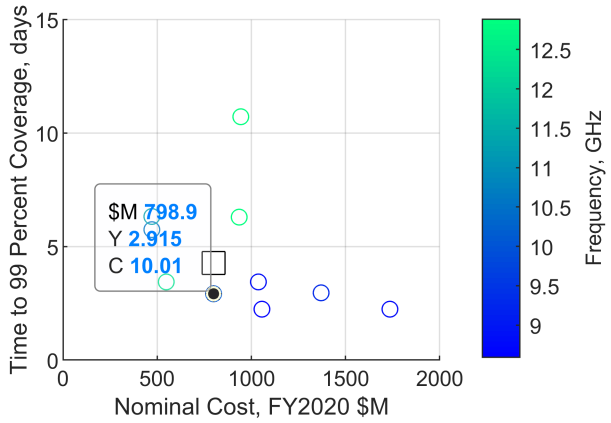


Figure 6.26: Downselected nondominated front for cost versus time to coverage.

transmitter (VV), horizontal transmitter (HH), downlink communications unit (Dwn), microwave radiometer (Mic), and nadir altimeter (Ndr). The final columns shown the estimated nonrecurring costs in millions of 2020 dollars (NRC), recurring costs in millions of 2020 dollars (RC), dry mass in kg (M_{dry}), and wet mass in kg (M_{wet}). Because of the low spacecraft masses, SSCM was used for cost prediction. NICM was used for cost prediction for the radar payload. Exactly one each of the VV transmitter, HH transmitter, nadir altimeter, and microwave radiometer are included in each formation. For the current optimization formulation, there is no benefit to having additional altimeters or radiometers.

Additional transmitters only decrease the average power used and therefore the size of the power subsystem, which for the given solution did not offset the cost of including additional transmitters. If the optimization is changed to instead illuminate a single ground area with multiple transmitters simultaneously, it is expected that the decrease in peak power and therefore transmitter mass will make the inclusion of multiple transmitters worthwhile. Two downlink units decrease the amount of data sent down by individual satellites, reducing the size of the communications subsystem. One downlink unit is placed on each satellite with a transmitter. This configuration is reasonable from a resilience standpoint because the loss of a transmitter degrades the performance of the formation and also reduces the amount of data downlinked. If a transmitter were lost without the loss of a downlink system, the downlink system would provide excess capacity. Therefore, no additional resilience is lost when including the downlink system with a transmitter. Because the downlink units are sized such that all downlink units are needed to transmit data, it is not beneficial to introduce additional downlink subsystems beyond the number of transmitters. If the downlink systems were designed to provide redundancy such that they could transmit sufficient data when one or more downlink systems are lost, additional units may prove beneficial. Because the radar transmitters are each part of a transmitter-downlink satellite, they are assumed to have the same configuration and therefore share nonrecurring costs.

Table 6.10: Selected solution hardware.

Sat Type	Num	VV	HH	Dwn	Mic	Ndr	NRC	RC	M_{dry}	M_{wet}
1	7	0	0	0	0	0	0.5	13.5	38.1	44.7
2	1	1	0	1	0	0	131.9	99.9	163.6	190.8
3	1	0	0	0	1	1	2.1	65.1	291.1	341.0
4	1	0	1	1	0	0	0	99.9	163.6	190.8

In addition to the three satellites with assigned payloads, there are seven satellites that have only receivers on board. The multiple receivers increase baseline diversity and prevent temporary configurations with no perpendicular baseline.

System properties and objective values are shown in Table 6.11. The radar frequency is in the X band. The solution includes two formations of ten satellites each. The average radar transmit power is of the solution is 26.1 W, similar to SWOT's average transmit power of 29.8 W. The peak power is driven down by the relationship between radar mass and peak power identified in the radar sizing analysis in subsection 3.4.2 and because of the correlation between peak radar power and cost in the NICM model. The radar antenna gain of 35.3 dBi is less than SWOT's gain of 50.5 dBi, but the path is significantly shorter due to the lower altitude of 430 km. Additionally, the lower frequency of the proposed solution reduces the power requirements. The maximum distance between the center of the formation and the edge of the formation in the radial direction is 3,627.2 m. The maximum distance between the center of the formation and the edge of the formation in the cross-track direction is 2.2 m. The launch costs are estimated at $24.5M. The launch manifest permits the selection of current or planned vehicles as of Fall 2019. If planned vehicles do not become available, it assumed that vehicles of comparable cost and capability will exist. The launch manifest required three vehicles, two Terran rockets and one Vector-R rocket. The first Terran rocket carries one radar satellite, one radiometer/altimeter satellite, and ten receive-only satellites. The second Terran rocket carries three radar satellites, one radiometer/altimeter satellite, and one receive-only satellite. The Vector-R rocket carries three receive-only satellites.

The required downlink rate is estimated at 490.3 Mbps per satellite with a downlink subsystem, for a total of 1,960.6 Mbps. The bandwidth per satellite is estimated to be 392 MHz. The two formations can be spatially separated so they are not simultaneously communicating with the same ground station to achieve a downlink rate of 980.3 Mbps per formation. Since the time to achieve coverage is 1/7 of the SWOT mission requirement of 21 days, the data burden can be reduced by sampling areas that are repeatedly covered less frequently. The SWOT mission plans a downlink rate of 610 Mbps. Reducing the sampling rate to SWOT's

Table 6.11: Selected solution parameters.

Parameter	Value
Antenna cross-track half-power beamwidth (°)	17.9
Antenna cross-track boresight angle (°)	11.9
Radar peak transmit power (W)	105.3
Altitude (km)	430.0
Inclination (°)	79.7
Radar center frequency (GHz)	10.0
Formation cross-track distance from center (m)	2.2
Formation radial distance from center (m)	3,627.2
Pulse repetition frequency (Hz)	3,191.2
Pulse width (μs)	77.8
Days until repeat	21
Number of formations	2
Radar transmitter mass (kg)	26.0
Downlink comms mass (kg)	33.1
Downlink transmit power (W)	1
Height error (cm)	7.3
Nominal cost (FY2020 $M)	798.9
Launch costs (FY2020 $M)	24.5
Time to 99% coverage (days)	2.9
Resilience	0.82

level would reduce the downlink rate per formation to 140 Mbps. The extra margin in the data rate occurs due to the fact that SWOT transmits higher resolution data for measurements over land, increasing the data burden in those areas. This increase is not considered in this analysis. The link budget is shown in Table 6.12 for a data rate of 490.3 Mbps and a bit error rate of 10^{-9}.

These results are compared to a cost analysis of the planned SWOT satellite. The SWOT program cost allocation is $1.15B according to a 2018 audit [5]. The proposed solution cost predicted by SSCM/NICM offers a 30% savings compared to the budgeted SWOT cost. However, the DISCO-Tech cost modelling analysis predicts the cost of the monolithic SWOT to be approximately $2B dollars using USCM/NICM, which is higher than the $1.15B allocation. USCM tends to provide higher cost estimates than SSCM, so the proposed solution

Table 6.12: Selected solution downlink budget.

Parameter	dB Value	Supporting Value
Downlink frequency		8 GHz
Transmit power	-0.54 dBW	0.83 W
Peak transmit antenna gain	15 dBi	
Transmit pointing loss	-0 dB	
Transmit line loss	-2.50 dB	
EIRP	11.96 dBW	
Space loss	-174.14 dB	
Atmospheric loss	-0.72 dB	
Propagation and polarization loss	-3.75 dB	
Modem implementation loss	-1.20 dB	
Losses	-179.81 dB	
Peak receive antenna gain	57.34 dBi	
Receive pointing loss	-0 dB	
Receiver line loss	-2 dB	
Receiver gain	55.34 dBi	
Boltzmann's constant, k		1.38×10^{-23} J/K
1/k	228.60 dBK/J	7.2×10^{22} K/J
Noise temperature, T		156.570 K
1/T	-21.947 dB/K	6.4×10^{-3} 1/K
Data rate, R		490.3 Mbps
1/R	-86.904 dB/Hz	2.04×10^{-9} s/bit
1/(kRT)	119.75 dB/Hz/J	
$\mathbf{E_b/N_0}$	7.24 dB	
Required $\mathbf{E_b/N_0}$	6.19 dB	
Link margin	1.05 dB	

is evaluated using USCM/NICM as well for consistency in modelling. The proposed solution cost estimated by USCM/NICM is \$1.44B. When using the same modelling techniques, the proposed solution still has an expected cost savings of 30% compared to the USCM/NICM-estimated SWOT cost. This cost savings is largely due to the decrease in radar peak power and the increase in pulse width.

6.2.5 Conclusions

This analysis identified several trends in disaggregated radar systems. Increasing the resilience of the mission requires an increase in number of satellites or satellite capability and therefore an increase in cost. Increasing the performance of the satellites also requires an increase in cost. Because more massive satellites have lower failure rates historically, increasing the complexity (represented by the mass) of critical satellites increases the resilience. However, a decrease in the payload complexity compared to the monolithic system can be achieved by dividing the radar components among multiple systems. This act allows the use of simpler, less massive, less costly satellites compared to the monolithic satellite. Cost decreases with increasing disaggregation, though the number of formations and the payload power have a more significant impact on cost than the disaggregation scheme. Increasing the level of disaggregation reduces the resilience to failure but increases the resilience to threat. The disaggregated formation also shows increased resilience to budget changes and cost overruns, since a partial formation can be manifested. The reduced complexity of the disaggregated satellites make them attractive alternatives in a field where complex configurations have historically resulted in cost overruns, schedule delay, and program cancellation. The disaggregated satellites are able to achieve sufficient performance to achieve the mission of the monolithic satellite.

Simulations of disaggregated solutions show that disaggregation could be a promising avenue for reducing mission cost and complexity for a SWOT-like SSH detection mission. Further study is needed to see if a disaggregated formation of satellites can provide similar performance for surface water level detection, SWOT's other objective. Surface water exhibits different backscattering than the ocean. The volumetric decorrelation is likely to be lower due to the lack of large waves, but the presence of land nearby can complicate the signal.

The ability to optimize complex, multistatic synthetic aperture radar systems represents a significant contribution to the field of disaggregated space system design.

Additionally, the methodology presented for evaluating trade-offs between cost and performance objectives is extensible to other multi-objective optimization problems. The contour surface analysis allows analysts to predict how sacrificing performance in one objective can improve performance in another. It also can be used to estimate the impact of budgetary changes on system performance. This analysis is therefore an important contribution to the field of space policy and decision making.

Chapter 7

Conclusions

This chapter summarizes conclusions from the scenarios examined throughout this dissertation, identifies significant contributions of the research, and provides recommendations for avenues of future research in the field of disaggregated space system design.

7.1 Conclusions

Conclusions will be discussed for the research questions identified in Chapter 1 and for the individual scenarios discussed throughout this dissertation. The overarching objective of this research project is to develop a methodology for disaggregation optimization that is broadly applicable and addresses gaps in capability for constellation and spacecraft design. This objective was achieved by the creation of a modular technique for optimizing constellation disaggregation problems using a genetic algorithm and evaluating performance through the use of cost, sizing, coverage, and other performance assessment modules. The DISCO-Tech methodology also allows the three research questions presented in Chapter 1 to be answered. The first research question was: How can multiple levels of disaggregation be easily represented and optimized? Through the validation scenarios presented in Chapter 4, it was shown that a variable-length genetic algorithm is a suitable choice for optimizing payload parameters, payload distribution, orbital parameters, and level of disaggregation. It was also shown that the variable-length algorithm outperforms the flagging method for varying

the number of components represented by a genome. This method allows the exploration of the design space without introducing biases through the selection of predefined architectures. The second research question was: How can resilience be assessed in a tractable and comprehensive manner? Chapter 5 presented a novel formulation for evaluating resilience to threats or resilience to worst-case failure for coverage-based analyses using linear programming techniques. This approach was applied to constellation design in the rideshare-based fire detection scenario and to the problem of ground station placement in the GPS ground station expansion scenario. Chapter 6 used an alternate formulation to assess resilience for a disaggregated system of satellites flying in formation. The third research question was: What are the trade-offs between cost, performance, resilience, and complexity for disaggregated and non-traditional systems? The trade-offs are dependent on the problem but generally showed that increasing cost allowed for increased numbers of satellites or increased payload capability, increasing performance. Resilience to failure was shown to increase with increasing disaggregation in the multi-orbit imaging scenario but was shown to decrease with increasing disaggregation in the interferometric SAR scenario. Resilience to threat was shown to increase with increasing disaggregation for both cases. Additionally, it was shown that disaggregation sometimes permitted the use of multiple less complex satellites in place of a monolithic satellite, reducing mission complexity and cost.

Through the rideshare-based fire detection analysis, it was shown that nanosatellites using commercial, off the shelf components could provide fire monitoring services with low revisit times for as little as $10.5 million. By leveraging rideshare opportunities, launch costs are greatly reduced, offsetting the cost of the additional satellites required to achieve the necessary coverage. Such a low cost constellation offers large returns on investments when compared to the annual costs of hundreds of billions of dollars due to fire damage in California. It was also shown that increasing the number of satellites in the imaging constel-

lations increased resilience to expected failure, since the likelihood that satellites providing successive coverage to a region will be lost is decreased. Resilience to threat increased with increasing numbers of satellites as well. Resilience to worst-case failure increased with small numbers of satellites but stagnated for large numbers of satellites.

DISCO-Tech was applied to the problem of determining the placement of additional monitoring stations for the GPS network. The analysis was focused on resilience, and it was desired that the ground stations provide continuous monitoring of the GPS satellite even if four ground stations are disabled. It was found that the addition of nine ground stations allows for continuous monitoring in the degraded case, bringing the total number of monitoring stations up to 26.

Finally, the performance of a disaggregated interferometric synthetic aperture radar formation was assessed. The formation was designed for the purpose of monitoring sea surface height. Sources of error were identified in relation to the formation configuration, number of satellites, and radar parameters. An analysis was performed that optimized the payload distribution among the satellites, orbital parameters, the relative spacing of the satellites, and radar parameters such as frequency and power. Objectives of measurement error, cost, coverage, and resilience to failure were assessed. A set of solutions were identified. Relationships between cost and performance were quantified to assess the potential performance gains for an increase in cost. It was shown that there are trade-offs between error and coverage for a fixed cost. For the sea surface height analysis, resilience is dependent only on the disaggregation scheme and therefore the cost. It was shown that increasing disaggregation raised a system's resilience to threat. If there were no redundant payloads, disaggregation decreased resilience to expected failure but increased the likelihood that the system would have some performance in the event of a failure. The inclusion of redundant payloads on disaggregated systems increased resilience to expected failure but at high cost compared to

monolithic systems. One solution was chosen from the set for further analysis. This solution offered cost savings of 30% above the reference monolithic satellite due to the simplification of payload components. The solution had improved coverage characteristics compared to the reference program and had an error of less than ten centimeters.

7.2 Significance of contributions

This research offers several significant contributions. The first contribution is a methodology for optimizing disaggregated systems that simultaneously optimizes payload distribution, level of disaggregation, payload parameters, and orbital parameters. Previous methodologies either varied payload distribution of fixed payloads for fractionated systems or varied payload parameters and orbital parameters for predefined architectures. DISCO-Tech uses a variable-length genetic algorithm to modify the number of satellite configurations or planes while simultaneously optimizing payload and orbital parameters. Because architectures are set by the algorithm, biases towards particular architectures are removed.

An additional key contribution of this research is a formulation for assessing resilience to threats for coverage-based constellations. The mixed integer linear programming approach developed exhaustively explores the threat space to determine the most damaging possible situation for a constellation. For large constellations and large numbers of satellite losses, the mixed integer linear program permits the solution of problems that would be intractable using full enumeration of the decision space. Formulations were presented for multiple revisit metrics including maximum revisit time and average time average gap. This formulation represents a key capability for designing defense systems that are resilient to adversarial action or for designing commercial systems that are resilient to system failures.

Another contribution of this research is the analysis of the cost and performance of a dis-

aggregated interferometric synthetic aperture radar system. The analysis examines error performance for the system and generates a set of solutions with a range of costs and capabilities. This detailed analysis quantifies the benefits of disaggregating interferometric synthetic aperture radar systems and provides a framework that can be used with other objectives or with newly developed commercial small satellite radar systems.

7.3 Recommendations for future research

The DISCO-Tech methodology has thus far been applied to constellation design problems and to ground station network design problems. The methodology can be extended to other domains, such as high-altitude balloon missions, unmanned aerial vehicle missions, or submarine missions. Additionally, hybrid solutions involving vehicles across one or more domains could be optimized using the DISCO-Tech framework. Future research could also assess the viability of disaggregated architectures for different types of space missions, such as agricultural monitoring, crime detection, debris removal, space-based computing, or deep-space optics. DISCO-Tech can also be incorporated into grant-writing processes to improve the fidelity and detail of proposals.

Because DISCO-Tech is modular, additional modules can be developed for different sensor types or objectives. A proposed module is a network resilience module, which assesses the ability of a constellation to reroute data flow when a satellite is lost. Specialized cost modules could be developed for nanosatellites and for hosted payloads to improve the fidelity of DISCO-Tech in those domains. A risk module that estimates the likelihood of mission success, such as the Complexity-Based Risk Assessment method, could also be implemented [14]. Additional sizing modules, such as an ADCS module that estimates mass based on spacecraft mass and pointing requirements or a data storage module that estimates mass

based on data load, could also be developed. The fidelity of the failure rate prediction model could be improved by incorporating estimates for component-level failure and launch failure rates. Additionally, future research could examine the impact of changes in input parameters such as failure rates on the nondominated front produced by the optimizer.

Future research could also include improvements to the optimizer. The genetic algorithm could be improved by the incorporation of domain-specific heuristics. Research could be done to identify a rigorous process for intelligently setting optimizer parameters such as mutation rate, length change rate, and initial population size. A methodology could also be developed to automatically determine when the use of alternative optimizers, such as interior point methods or simplex methods, is beneficial. Other optimizers that could be investigated include the DIviding RECTangles (DIRECT) method and stochastic approximation techniques.

Bibliography

[1] Mission to planet earth: Topex/poseidon. Technical Report 92-126, NASA/CNES, 1992.

[2] Attenuation by atmospheric gases. Technical Report ITU-R Recommendation P.676-9, International Telecommunication Union, 2012.

[3] Swot: The surface water and ocean topography mission - wide-swath altimetric measurement of water elevation on earth. Technical report, NASA Jet Propulsion Laboratory, 2012.

[4] Swot project mission performance and error budget. Technical Report D-79084, NASA Jet Propulsion Laboratory, 2017.

[5] Nasa's surface water and ocean topography mission. Technical Report IG-18-011, NASA Office of Inspector General, 2018.

[6] Ossama Abdelkhalik and Daniele Mortari. Orbit design for ground surveillance using genetic algorithms. *Journal of Guidance, Control, and Dynamics*, 29(5):1231–1235, 2006. doi: 10.2514/1.16722.

[7] B Acikmese, JM Carson, and L Blackmore. Lossless convexification of the soft landing optimal control problem with non-convex control bound and pointing constraints. *IEEE Transactions on Control Systems Technology*, 21(6):2104–2113, 2013.

[8] Salvatore Alfano. Rapid generation of site/satellite timetables. *Journal of Spacecraft and Rockets*, 30(6):760–764, 1993. doi: 10.2514/3.26383.

[9] Mark Andraschko, Jeffrey Antol, Stephen Horan, and Doreen Nei. Commercially hosted government payloads: Lessons from recent programs. In *2011 IEEE Aerospace Conference*, pages 1–15, Big Sky, MT, 2011. Institute of Electrical and Electronics Engineers. doi: 10.1109/AERO.2011.5747491.

[10] Leonid Appel, Moshe Guelman, and David Mishne. Optimization of satellite constellation reconfiguration maneuvers. *Acta Astronautica*, 99:166–174, Jun 2014. doi: 10.1016/j.actaastro.2014.02.016.

[11] Gautam D Badhwar and Phillip D Anz-Meador. Determination of the area and mass distribution of orbital debris fragments. *Earth, Moon, and Planets*, 45(1):29–51, 1989.

[12] G Balmino, B Moynot, and N Vales. Gravity field model of mars in spherical harmonics up to degree and order eighteen. *Journal of Geophysical Research: Solid Earth*, 87 (B12):9735–9746, 1982.

[13] David J Barnhart, Tanya Vladimirova, and Martin N Sweeting. Very-small-satellite design for distributed space missions. *Journal of Spacecraft and Rockets*, 44(6):1294–1306, 2007. doi: https://doi.org/10.2514/1.28678.

[14] David Bearden, Mark Cowdin, and Justin Yoshida. Evolution of complexity and cost for planetary missions throughout the development lifecycle. In *2012 IEEE Aerospace Conference*, pages 1–12. IEEE, 2012.

[15] K. L. Bedingfield, R. D. Leach, and M. B. Alexander. Spacecraft system failures and anomalies attributed to the natural space environment. Technical Report 1390, NASA, 1996.

[16] J. Berk, J. Straub, and D. Whalen. The open prototype for educational nanosats:

Fixing the other side of the small satellite cost equation. In *2013 IEEE Aerospace Conference*, pages 1–16, March 2013. doi: 10.1109/AERO.2013.6497393.

[17] Willy Bertiger, Shailen D Desai, Angie Dorsey, Bruce J Haines, Nate Harvey, Da Kuang, Ant Sibthorpe, and Jan P Weiss. Sub-centimeter precision orbit determination with gps for ocean altimetry. *Marine Geodesy*, 33(S1):363–378, 2010. doi: 10.1080/01490419.2010.487800.

[18] Dimitri P Bertsekas. Convexification procedures and decomposition methods for non-convex optimization problems. *Journal of Optimization Theory and Applications*, 29 (2):169–197, 1979.

[19] Louis Breger and Jonathan P. How. Gve-based dynamics and control for formation flying spacecraft. In *Proceedings of the 2nd International Symposium on Formation Flying Missions & Technologies*, Washington, D.C., 2004. NASA.

[20] Mike Brown. Spacex has a bold timeline for getting to mars and starting a colony. *Inverse*, 2019.

[21] Stefan Buckreuss and Manfred Zink. The missions terrasar-x and tandem-x: Status, challenges, future perspectives. page 1, 2011. doi: 10.1109/URSIGASS.2011.6050872.

[22] Ron Burch. A method for calculation of the resilience of a space system. In *MILCOM 2013-2013 IEEE Military Communications Conference*, pages 1002–1007. IEEE, 2013.

[23] Pau Garcia Buzzi, Daniel Selva, Nozomi Hitomi, and William J Blackwell. Assessment of constellation designs for earth observation: Application to the tropics mission. *Acta Astronautica*, 161:166–182, 2019.

[24] Julie C. Castillo-Rogez, Karen Meech, Soon-Jo Chung, and Damon Landau. Approach to exploring interstellar objects and long-period comets. In *Advances in the Astronau-*

tical Sciences AAS/AIAA Space Flight Mechanics Meeting 2019. Univelt, 2019. Pre publication.

[25] Jing Chu, Jian Guo, and Eberhard KA Gill. Novel approaches to the design of fractionated clusters for long-term earth observation missions. *idea*, 3:4, 2013.

[26] Jing Chu, Jian Guo, and Eberhard Gill. Functional, physical, and organizational architectures of a fractionated space infrastructure for long-term earth observation missions. *IEEE Aerospace and Electronic Systems Magazine*, 29(12):6–17, 2014.

[27] Maurice Clerc. Discrete particle swarm optimization, illustrated by the traveling salesman problem. In *New optimization techniques in engineering*, pages 219–239. Springer, 2004.

[28] John M. Colombi, Lyndsey D. Buckle, Jonathan T. Black, and Sarah G. Nurre. Optimal launch manifesting for heterogeneous disaggregated satellite constellations. *Journal of Spacecraft and Rockets*, 54(3):582–591, 2017. doi: 10.2514/1.A33796.

[29] John M. Colombi, Jordan L. Stern, Steven T. Wachtel, David W. Meyer, and Richard G. Cobb. Multi-objective parallel optimization of geosynchronous space situational awareness architectures. *Journal of Spacecraft and Rockets*, 55(6):1453–1465, 2018. doi: 10.2514/1.A34043.

[30] Air Force Space Command. Resiliency and disaggregated space architectures. Technical Report AFD-130821-034, Air Force Space Command, 2013.

[31] Sarah W. Cooley, Laurence C. Smith, Leon Stepan, and Joseph Mascaro. Tracking dynamic northern surface water changes with high-frequency planet cubesat imagery. *Remote Sensing*, 9(12), 2017. doi: 10.3390/rs9121306.

[32] National Research Council. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. The National Academies Press, Washington, DC, 2007. ISBN 978-0-309-14090-4. doi: 10.17226/11820.

[33] B. G. W. Craenen, A. E. Eiben, and E. Marchiori. *Practical Handbook Of Genetic Algorithms: Applications*, chapter How to handle constraints with evolutionary algorithms., pages 341–361. CRC Press, Boca Raton, FL, second edition, 2001.

[34] William A Crossley and Edwin A Williams. Simulated annealing and genetic algorithm approaches for discontinuous coverage satellite constellation design. *Engineering Optimization*, 32(3):353–371, 2000.

[35] Joseph J Cuhran, Mathew K Jenkins, and Michael J Walters. Quantifying and evaluating the resilience of optimized space constellations for fire detection. Master's thesis, AIR FORCE INSTITUTE OF TECHNOLOGY WRIGHT-PATTERSON AFB, 2017.

[36] John C Curlander and Robert N McDonough. *Synthetic aperture radar- Systems and signal processing*. John Wiley & Sons, Inc, 1991.

[37] Simone D'Amico, Oliver Montenbruck, Christian Arbinger, and Hauke Fiedler. Formation flying concept for close remote sensing satellites. In *15th AAS/AIAA Space Flight Mechanics Conference*, pages 05–156. AAS Washington, DC, 2005.

[38] Simone D'Amico, Marco Pavone, Shailendhar Saraf, Abdulaziz Alhussien, Turki Al-Saud, Sasha Buchman, Robert Byer, and Charbel Farhat. Miniaturized autonomous distributed space system for future science and exploration. In *International Workshop on Satellite Constellations and Formation Flying*, pages 1–20. International Astronautical Federation's Astrodynamics Committee Delft, The ..., 2015.

[39] Jeremy J. Davis. *Constellation reconfiguration: Tools and analysis*. PhD thesis, Texas A&M University, College Station, TX, 2010.

[40] Jeremy J. Davis, Martín E. Avendaño, and Daniele Mortari. The 3-d lattice theory of flower constellations. *Celestial Mechanics and Dynamical Astronomy*, 116(4):339–356, 2013. doi: 10.1007/s10569-013-9494-7.

[41] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elistist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. doi: 10.1109/4235.996017.

[42] Marco D&'Errico. *Distributed space missions for earth system monitoring*, volume 31. Springer Science & Business Media, 2012.

[43] Shubhangi Deshpande, Layne T. Watson, and Robert A. Canfield. Multiobjective optimization using an adaptive weighting scheme. *Optimization Methods and Software*, 31(1), 2016. doi: https://doi.org/10.1080/10556788.2015.1048861.

[44] William J Donnelly, James R Carswell, Robert E McIntosh, Paul S Chang, John Wilkerson, Frank Marks, and Peter G Black. Revised ocean backscatter models at c and ku band under high-wind conditions. *Journal of Geophysical Research: Oceans*, 104(C5):11485–11497, 1999.

[45] Gregory F Dubos and Joseph H Saleh. Comparative cost and utility analysis of monolith and fractionated spacecraft using failure and replacement markov models. *Acta Astronautica*, 68(1-2):172–184, 2011.

[46] Gregory F. Dubos, Jean-Francois Castet, and Joseph H. Saleh. Statistical reliability analysis of satellites by mass category: Does spacecraft size matter? *Acta Astronautica*, 67:584–595, 2010. doi: 10.1016/j.actaastro.2010.04.017.

[47] Dynetics. The gremlins program fact sheet. Technical Report K-170437, Dynetics, 2017. https://www.dynetics.com/_files/strike-systems/Dynetics%20Gremlins.pdf.

[48] Takuji Ebinuma, Robert H Bishop, and E Glenn Lightsey. Integrated hardware investigations of precision spacecraft rendezvous using the global positioning system. *Journal of guidance, control, and dynamics*, 26(3):425–433, 2003. doi: https://doi.org/10.2514/2.5080.

[49] Armin Ellis, Michael Mercury, and Shannon Brown. Global coverage from ad-hoc constellations in rideshare orbits. In *Proceedings of the 26th Annual AIAA/USU Conference on Small Satellites*, Logan, UT, 2012. The American Institute of Aeronautics and Astronautics.

[50] D. Escorial, I. F. Tourne, F. J. Reina, and J. Gonzalo. Fuego: A dedicated constellation of small satellites to detect and monitor forest fires. *Acta Astronautica*, 53(9-12):765–775, 2003. doi: 10.1016/S0094-5765(03)00052-3.

[51] Madhi Fakoor, Majid Bakhtiari, and Mahshid Soleymani. Optimal design of the satellite constellation arrangement reconfiguration process. *Advances in Space Research*, 58(3):372–386, 2016. doi: 10.1016/j.asr.2016.04.031.

[52] Tom G. Farr and Mike Kobrick. The shuttle radar topography mission. Technical Report ADP010831, NASA Jet Propulsion Lab, 2000.

[53] M. P. Ferringer, D. B. Spencer, and P. Reed. Many-objective reconfiguration of operational satellite constellations with the large-cluster epsilon non-dominated sorting genetic algorithm-ii. In *2009 IEEE Congress on Evolutionary Computation*, pages 340–349, Trondheim, Norway, May 2009. Institute of Electrical and Electronics Engineers. doi: 10.1109/CEC.2009.4982967.

[54] Matthew Ferringer, Marc DiPrinzio, Timothy Thompson, Kyle Hanifen, and Patrick Reed. A framework for the discovery of passive-control, minimum energy satellite constellations. In *AIAA/AAS Astrodynamics Specialist Conference*, page 4158, 2014.

[55] Matthew P Ferringer and David B Spencer. Satellite constellation design tradeoffs using multiple-objective evolutionary computation. *Journal of spacecraft and rockets*, 43(6):1404–1411, 2006.

[56] Matthew P Ferringer, Ronald S Clifton, and Timothy G Thompson. Efficient and accurate evolutionary multi-objective optimization paradigms for satellite constellation design. *Journal of Spacecraft and Rockets*, 44(3):682–691, 2007. doi: 10.2514/1.26747.

[57] Mark Fleischer. The measure of pareto optima applications to multi-objective metaheuristics. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 519–533. Springer, 2003.

[58] Karen Fletcher, editor. *InSAR processing: a mathematical approach*, pages C–111. ESA Publications, 2007. ISBN: 92-9092-233-8.

[59] Veronica Foreman, Jacqueline Le Moigne, and Olivier de Weck. Satellite constellation cost modeling: An aggregate model. Technical Report 20180007289, NASA Earth Science Technology Office, 2016.

[60] Anders Forsgren, Philip E Gill, and Margaret H Wright. Interior methods for nonlinear optimization. *SIAM review*, 44(4):525–597, 2002.

[61] Joseph W. Gangestad, James R. Wilson, Kristin L. Gates, and John V. Langer. Rideshare-initiated constellations: Future cubesat architectures with the current

launch manifest. In *Proceedings of the 31st Space Symposium*, Colorado Springs, CO, 2015. Space Foundation, Space Foundation.

[62] W. H. Gerstenmaier. Progress in defining the deep space gateway and transport plan. Technical report, NASA, March 2017.

[63] Alessandro Golkar and Ignasi Lluch i Cruz. The federated satellite systems paradigm: Concept and business case evaluation. *Acta Astronautica*, 111:230–248, Jun 2015. doi: 10.1016/j.actaastro.2015.02.009.

[64] Donald E Grierson. Pareto multi-criteria decision making. *Advanced Engineering Informatics*, 22(3):371–384, 2008.

[65] John G. Grimes. Global positioning system standard positioning service performance standard. Technical Report GPS SPS PS, Department of Defense, 2008.

[66] Cesare Guariniello, Melanie Grande, Christopher Brand, Liam Durbin, Michael Dai, Ashwati Das-Stuart, Reginald A Alexander, Kathleen Howell, and Daniel DeLaurentis. Quantifying the impact of systems interdependencies in space systems architectures. 2019.

[67] Cesare Guariniello, Linas Mockus, Ali K Raz, and Daniel A DeLaurentis. Towards intelligent architecting of aerospace system-of-systems. In *2019 IEEE Aerospace Conference*, pages 1–11. IEEE, 2019.

[68] David Hadka and Patrick Reed. Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computing*, 21(2):231–259, 2013. doi: 10.1162/EVCO_a_00075.

[69] Daniel E. Hastings and Paul A. La Tour. An economic analysis of disaggregation of

space assets: Application to gps. *Acta Astronautica*, 134:244–264, May 2017. doi: 10.1016/j.actaastro.2017.02.008.

[70] Daniel E. Hastings, Paul A. La Tour, and Dipak Kumar Giri. Architectural concept of disaggregation of space systems: application to weather satellites. *Journal of Spacecraft and Rockets*, 2019. doi: 10.2514/1.A34353. Pre-publication.

[71] R. L. Haupt and S. E. Haupt. *Practical Genetic Algorithms*. Wiley and Sons, Hoboken, NJ, second edition, 2004.

[72] Hank Heidt, Jordi Puig-Suari, Augustus Moore, Shinichi Nakasuka, and Robert Twiggs. Cubesat: A new generation of picosatellite for education and industry low-cost space experimentation. 2000.

[73] Nozomi Hitomi and Daniel Selva. Constellation optimization using an evolutionary algorithm with a variable-length chromosome. In *2018 IEEE Aerospace Conference*, pages 1–12. IEEE, 2018.

[74] Nozomi Hitomi, Hyunseung Bang, and Daniel Selva. Adaptive knowledge-driven optimization for architecting a distributed satellite system. *Journal of Aerospace Information Systems*, 15(8):485–500, 2018.

[75] Bernhard Hofmann-Wellenhof, Herbert Lichtenegger, and Elmar Wasle. *GNSS – Global Navigation Satellite Systems*, chapter GPS, pages 309–340. Springer, 2008.

[76] John H Holland. Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. 1975.

[77] Marcus J. Holzinger, Daniel J. Scheeres, and R. Scott Erwin. On-orbit operational range computation using gauss's variational equations with j2 perturbations. *Journal of Guidance, Control, and Dynamics*, 37(2):608–622, 2014. doi: 10.2514/1.53861.

[78] Baisravan HomChaudhuri, Meeko Oishi, Matt Shubert, Morgan Baldwin, and R. Scott Erwin. Computing reach-avoid sets for space vehicle docking under continuous thrust. In *2016 IEEE 55th Conference on Decision and Control*, pages 3312–3318, Las Vegas, NV, 2016. Institute of Electrical and Electronics Engineers. doi: 10.1109/CDC.2016.7798767.

[79] Lester Ingber. Adaptive simulated annealing (asa): Lessons learned. *Polish Journal Control and Cybernetics, Special Edition: Simulated Annealing Applied to Combinatorial Optimization*, 1995.

[80] Ch Jayles, JP Chauveau, and F Rozo. Doris/jason-2: better than 10 cm on-board orbits available for near-real-time altimetry. *Advances in space research*, 46(12): 1497–1512, 2010. doi: 10.1016/j.asr.2010.04.030.

[81] D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Application*, 79(1), 1993.

[82] Patrick Kelly and Riccardo Bevilacqua. The constellation for mars position acquisition using small satellites: Cubesat design feasibility and challenges. In *Advances in the Astronautical Sciences 4th IAA Dynamics and Control of Space Systems Conference*, volume 163, pages 629–640. Univelt, 2017.

[83] James Kennedy and Russell Eberhart. Particle swarm optimization (pso). In *Proceedings of the IEEE International Conference on Neural Networks*, pages 1942–1948, 1995.

[84] Vincent Kerbaol, Bertrand Chapron, and Paris W. Vachon. Analysis of ers-1/2 synthetic aperture radar wave mode imagettes. *Journal of Geophysical Research*, 103 (C4):7833–7846, 1998.

[85] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[86] Joshua Knowles and David Corne. On metrics for comparing nondominated sets. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, volume 1, pages 711–716. IEEE, 2002. doi: 10.1109/CEC.2002.1007013.

[87] Thorsten Koch, Tobias Achterberg, Erling Andersen, Oliver Bastert, Timo Berthold, Robert E. Bixby, Emilie Danna, Gerald Gamrath, Ambros M. Gleixner, Stefan Heinz, Andrea Lodi, Hans Mittelmann, Ted Ralphs, Domenico Salvagnin, Daniel E. Steffy, and Kati Wolter. Miplib 2010. *Mathematical Programming Computation*, 3(2): 103, Jun 2011. doi: 10.1007/s12532-011-0025-9.

[88] Adam W Koenig, Simone D'Amico, Bruce Macintosh, and Charles J Titus. A pareto-optimal characterization of miniaturized distributed occulter/telescope systems. In *Techniques and Instrumentation for Detection of Exoplanets VII*, volume 9605, page 960514. International Society for Optics and Photonics, 2015.

[89] Adam W Koenig, Bruce Macintosh, and Simone D'Amico. Formation design of distributed telescopes in earth orbit for astrophysics applications. *Journal of Spacecraft and Rockets*, pages 1–16, 2019.

[90] A Kohlhase, S D Amico, and R Kroes. Evaluating interferometric baseline performances in a close formation flight by using grace gps navigation solutions. In *PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON SPACE TECHNOLOGY AND SCIENCE*, volume 25, page 599, 2006.

[91] Joshua B. Kollat and Patrick M. Reed. The value of online adaptive search: a performance comparison of nsgaii, ϵ-nsgaii and ϵmoea. In *International Conference*

on Evolutionary Multi-Criterion Optimization, pages 386–398, Berlin, Hiedelberg, 2005. Springer.

[92] Jan Kolmas, Payam Banazadeh, Adam W Koenig, Bruce Macintosh, and Simone D'Amico. System design of a miniaturized distributed occulter/telescope for direct imaging of star vicinity. In *2016 IEEE Aerospace Conference*, pages 1–11. IEEE, 2016.

[93] John Daniel Kraus and Ronald J Marhefka. *Antennas for all applications*. 2002.

[94] Gerhard Krieger, Hauke Fiedler, Irena Hajnsek, Michael Eineder, Marian Werner, and Alberto Moreira. Tandem-x: mission concept and performance analysis. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 7, pages 4890–4893, 2005. doi: 10.1109/igarss.2005.1526770.

[95] Paul Alexis La Tour. *Combining tradespace exploration with system dynamics to explore future space architectures*. PhD thesis, Massachusetts Institute of Technology, 2016.

[96] Jarret M. Lafleur. Exploring the f6 fractionated spacecraft trade space with gt-fast. Master's thesis, Georgia Institute of Technology, 2009.

[97] Juan F. Lam, Paul La Tour, , and Jim Kim. Architecting disaggregated space weather solutions using nanosatellites. In *AIAA SPACE 2013 Conference and Exposition*, AIAA Space Forum. American Institute of Aeronautics and Astronautics, 2013. doi: 10.2514/6.2013-5322.

[98] Wiley J. Larson and James R. Wertz, editors. *Space Mission Analysis and Design*, chapter Communications Architecture, pages 533–586. Space technology library.

Microcosm Press and Kluwer Academic Publishers, El Segundo, CA, third edition, 1999.

[99] C. Y. Lee and E. K. Antonsson. Variable length genomes for evolutionary algorithms. In *Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation*, page 806, Las Vegas, NV, 2000. Association for Computing Machinery Special Interest Group on Genetic and Evolutionary Computation, Morgan Kaufmann Publishers Inc.

[100] Robert S. Legge, Jr and David W. Miller. *Optimization and Valuation of Reconfigurable Satellite Constellations Under Uncertainty*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2014.

[101] Ching-Jong Liao, Chao-Tang Tseng, and Pin Luarn. A discrete version of particle swarm optimization for flowshop scheduling problems. *Computers & Operations Research*, 34(10):3099–3111, 2007.

[102] Zhong Lu and Daniel Dzurisin. *InSAR Imaging of Aleutian Volcanoes: Monitoring a Volcanic Arc from Space*, chapter Introduction to Interferometric Synthetic Aperture Radar, pages 1–23. Springer-Verlag, 2014. doi: 10.1007/978-3-642-00348-6_1.

[103] Kurt Mann, Douglas Holker, and Nate Conn. Responsive environmental assessment commercially hosted (reach) payloads. Institute of Electrical and Electronics Engineers, 2017. doi: 10.1109/AERO.2017.7943761.

[104] Salvo Marcuccio, Pierpaolo Pergola, Stefan Gregucci, and Mariano Andrenucci. Low-thrust propulsion systems for small satellites. In *Proceedings of the 66th International Astronautical Congress*, Jerusalem, Israel, 2015. International Astronautical Federation.

[105] Matthew L Marcus and Raymond J Sedwick. Low earth orbit debris removal technology assessment using genetic algorithms. *Journal of Spacecraft and Rockets*, 54(5):1110–1126, 2017.

[106] Maurice Martin and Michael Stallard. Distributed satellite missions and technologies-the techsat 21 program. In *Space Technology Conference and Exposition*, page 4479, 1999.

[107] John L Maryak and Daniel C Chin. Global random optimization by simultaneous perturbation stochastic approximation. *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, 53(3):781, 2008.

[108] R. H. Merson. The motion of a satellite in an axi-symmetric gravitation field. *Geophysical Journal of the Royal Astronomical Society*, 4(17):111–128, 1961. doi: 10.1007/BF00643817.

[109] Josef Mittermayer, Paco Lopez-Dekker, Thomas Kraus, and Gerhard Krieger. Small satellite dispersed sar-an exemplary configuration. In *Proceedings of EUSAR 2016: 11th European Conference on Synthetic Aperture Radar*, pages 1–4. VDE, 2016.

[110] Jacqueline Le Moigne, Philip Dabney, Olivier de Weck, Veronica Foreman andPaul Grogan, Matthew Holland, Steven Hughes, and Sreeja Nag. Tradespace analysis tool for designing constellation (tat-c). In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1181–1184, Fort Worth, TX, 2017. IEEE.

[111] Ki mook Kang and Duk jin Kim. Feasibility study on estimating sea surface currents from single (envisat asar) and dual (tandem-x) sar system. pages 1274–1277, 2013. doi: 10.1109/IGARSS.2013.6723013.

[112] Rosemary Morrow, Lee-Lueng Fu, Fabrice Ardhuin, Mounir Benkiran, Bertrand Chapron, Emmanuel Cosme, Francesco d'Ovidio, J. Thomas Farrar, Sarah T. Gille, Guillaume Lapeyre, Pierre-Yves Le Traon, Ananda Pascual, Aurélien Ponte, Bo Qiu, Nicolas Rascle, Clement Ubelmann, Jinbo Wang, and Edward D. Zaron. Global observations of fine-scale ocean surface topography with the surface water and ocean topography (swot) mission. *Frontiers in Marine Science*, 6:232, 2019. ISSN 2296-7745. doi: 10.3389/fmars.2019.00232. URL https://www.frontiersin.org/article/10.3389/fmars.2019.00232.

[113] Daniele Mortari, Matthew P. Wilkins, and Christian Bruccoleri. The flower constellations. *Journal of Astronautical Sciences*, 52(1), 2004.

[114] Katherine E. Mott and Jonathan T. Black. Heterogeneous constellation design methodology applied to a mars-orbiting communications and positioning constellation. In *AAS/AIAA Astrodynamics Conference, 2017*, volume 162 of *Advances in the Astronautical Sciences Series*, pages 2383–2396, Stevenson, WA, 2017. Univelt.

[115] Katherine E. Mott and Jonathan T. Black. Model-based heterogeneous optimal space constellation design. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 602–609. Institute of Electrical and Electronics Engineers, Jul 2018. doi: 10.23919/ICIF.2018.8455222.

[116] Katherine E. Mott and Jonathan T. Black. Design of a resilient rideshare-based small satellite constellation using a genetic algorithm. In *Advances in the Astronautical Sciences AAS/AIAA Space Flight Mechanics Meeting 2019*. Univelt, 2019.

[117] Joel N. Myers. Accuweather predicts 2018 wildfires will cost california total economic losses of $400 billion. *AccuWeather*, 2019.

[118] Sreeja Nag, Steven P. Hughes, and Jacqueline J. Le Moigne. Streamlining the design tradespace for earth imaging constellations. In *AIAA SPACE 2016*, pages 1–17, Long Beach, CA, 2016. AIAA. doi: 10.2514/6.2016-5561.

[119] Sreeja Nag, Steven P Hughes, and Jacqueline J LeMoigne. Navigating the deployment and downlink tradespace for earth imaging constellations. In *Proceedings of the 66th International Astronautical Congress*. International Astronautical Federation, 2017.

[120] NASA. Elana xii cubesat launch on nrol-55 mission. Technical report, NASA, 2015.

[121] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

[122] Xin Ning, Jianping Yuan, and Xiaokui Yue. Uncertainty-based optimization algorithms in designing fractionated spacecraft. *Scientific Reports*, 6:1–8, 2016. doi: 10.1038/srep22979.

[123] Andrew Nuss. Hydra. Technical report, DARPA, 2018. https://www.darpa.mil/program/hydra.

[124] Michael Gregory O'Neill and Annalisa L. Weigel. Assessing fractionated spacecraft value propositions for earth imaging space missions. *Journal of Spacecraft and Rockets*, 48(6):974–986, 2011. doi: 10.2514/1.50947.

[125] Ellen Pawlikowski, Doug Loverro, and Tom Cristler. Space: Disruptive challenges, new opportunities, and new strategies. Technical report, Air Force Space Command, Los Angeles Air Force Base, CA, 2012.

[126] Douglas J. Pegher and Jason A. Parish. Optimizing coverage and revisit time in sparse military satellite constellations: a comparison of traditional approaches and genetic algorithms. Master's thesis, Naval Postgraduate School, Monterey, CA, 2004.

[127] Zong Peng and Saeid Kohani. The performance of the constellations satellites based on reliability. *Journal of Space Safety Engineering*, 4(2):112–116, 2017. doi: 10.1016/j.jsse.2017.07.003.

[128] Nam D. Pham. The economic benefits of commercial gps use in the u.s. and the costs of potential disruption. Technical report, NDP Consulting, 2011.

[129] David Mostaza Prieto, Benjamin P Graziano, and Peter CE Roberts. Spacecraft drag modelling. *Progress in Aerospace Sciences*, 64:56–65, 2014.

[130] Carol Rasmussen. 25 years of global sea level data, and counting. Technical Report 2017-212, NASA Jet Propulsion Laboratory, 2017.

[131] Paul Rosen, Michael Eineder, B Rabus, Eric Gurrola, Scott Hensley, Walter Knöpfle, H Breit, Achim Roth, and Marian Werner. Srtm-mission-cross comparison of x and c band data properties. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)*, volume 2, pages 751–753. IEEE, 2001.

[132] Matthew Lee Ryerkerk. *Metameric representations in evolutionary algorithms*. PhD thesis, Michigan State University, 2018.

[133] Daniel Selva. Knowledge-intensive global optimization of earth observing system architectures: a climate-centric case study. In *Sensors, Systems, and Next-Generation Satellites XVIII*, volume 9241, page 92411S. International Society for Optics and Photonics, 2014.

[134] Daniel Selva and Edward F Crawley. Vassar: Value assessment of system architectures using rules. In *2013 IEEE Aerospace Conference*, pages 1–21. IEEE, 2013.

[135] Randy Showstack. Sentinel satellites initiate new era in earth observation. *Eos, Transactions American Geophysical Union*, 95(26):239–240, 2014.

[136] Puneet Singla, Kamesh Subbarao, and John L Junkins. Adaptive output feedback control for spacecraft rendezvous and docking under measurement uncertainty. *Journal of guidance, control, and dynamics*, 29(4):892–902, 2006. doi: https://doi.org/10.2514/1.17498.

[137] Bryce Space and Technology. Smallsats by the numbers 2018. Technical report, Bryce Space and Technology, 2018.

[138] James C Spall et al. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE transactions on automatic control*, 37(3): 332–341, 1992.

[139] Philip N. Springmann and Olivier L. de Weck. Parametric scaling model for nongeosynchronous communications satellites. *Journal of Spacecraft and Rockets*, 41 (3):472–477, 2004. doi: 10.2514/1.6220.

[140] Detlef Stammer and Anny Cazenave. *Satellite altimetry over oceans and land surfaces*. CRC Press, 2017.

[141] Douglas K. Stenger. Survivability analysis of the iridium low earth orbit satellite network. Master's thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 1996.

[142] STK. The coverage figure of merit object. Online, 2016. https://help.agi.com/stk/11.0.1/Content/stk/ObjectMap_CoverageFOM.htm.

[143] STK. Measuring the average length of the coverage gap. Online, 2018. http://help.agi.com/stk/index.htm#cov/fom-10.htm.

[144] Cornel Sultan. Decoupling approximation design using the peak to peak gain. *Mechanical Systems and Signal Processing*, 36(2):582–603, 2013.

[145] Shashank Tamaskar and Daniel DeLaurentis. Modular spacecraft architecture, a new paradigm in spacecraft design. 2010.

[146] Shashank Tamaskar, Daniel DeLaurentis, and Kartavya Neema. Complexity analysis of spacecraft architectures. In *AIAA SPACE 2011 Conference & Exposition*, page 7201, 2011.

[147] Mohit Tawarmalani and Nikolaos V Sahinidis. *Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications*, volume 65. Springer Science & Business Media, 2013.

[148] Jim Taylor, editor. *Deep Space Communications*, chapter Mars Reconnaissance Orbiter, pages 201–259. JPL Deep-Space Communications and Navigation Series. John Wiley and Sons, Inc., Hoboken, NJ, first edition, 2016.

[149] Dylan Thomas, Katherine Mott, Kristen Tetreault, Kevin M Nastasi, Ian Elliott, Robert Scheible, Ethan Ohriner, and Jonathan Black. Real-time on-board estimation and optimal control of autonomous micro-satellite proximity operations. In *55th AIAA Aerospace Sciences Meeting, AIAA SciTech Forum*. The American Institute of Aeronautics and Astronautics, 2017. doi: 10.2514/6.2017-0398.

[150] R. E. Thompson, J. M. Colombi, J. Black, and B. J. Ayres. Disaggregated space system concept optimization: Model-based conceptual design methods. *Systems Engineering*, 18(6):549–567, 2015. doi: 10.1002/sys.21310.

[151] R. E. Thompson, J. M. Colombi, J. Black, and B. J. Ayres. Model-based conceptual

design optimization methods: Disaggregated weather system follow-on. *Journal of Spacecraft and Rockets*, 52(4):1021–1037, 2015. doi: 10.2514/1.A33135.

[152] Robert E. Thompson. *A Methodology for the Optimization of Disaggregated Space System Conceptual Designs*. PhD thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2015.

[153] Clement Ubelmann, Patrice Klein, and Lee-Lueng Fu. Dynamic interpolation of sea surface height and potential applications for future high-resolution altimetry mapping. *Journal of Atmospheric and Oceanic Technology*, 32:177–184, 2015. doi: 10.1175/JTECH-D-14-00152.1.

[154] D. A. Vallado. *Fundamentals of Astrodynamics and Applications*, chapter General Perturbation Techniques, pages 609–654. Microcosm Press, Hawthorne, CA, fourth edition, 2013.

[155] Catherine C. Venturini. Improving mission success of cubesats. Technical Report TOR-2017-01689, The Aerospace Corporation, 2017.

[156] Maria-José Viñas. Operation icebridge, icesat-2 join forces to survey antarctica. *NASA's Earth Science News*, 2018. URL https://www.nasa.gov/feature/goddard/2018/operation-icebridge-icesat-2-join-forces-to-survey-antarctica.

[157] Antoni Viros Martin and Daniel Selva. From design assistants to design peers: Turning daphne into an ai companion for mission designers. In *AIAA Scitech 2019 Forum*, page 0402, 2019.

[158] Katherine M. Wagner, Kevin K. Schroeder, and Jonathan T. Black. Disaggregated constellation optimization applied to environmental monitoring. In *Proceedings of the*

70th International Astronautical Congress. International Astronautical Federation, 2019.

[159] J. G. Walker. Circular orbit patterns providing continuous whole earth coverage. Technical Report 70211, Royal Aircraft Establishment, 1970.

[160] M. J. H. Walker. Errata, a set of modified equinoctial orbital elements. *Celestial Mechanics*, 38(4):391–392, 1986. doi: 10.1007/BF01238929.

[161] M. J. H. Walker, B. Ireland, and Joyce Owens. A set of modified equinoctial orbital elements. *Celestial Mechanics*, 36(4):409–419, 1985. doi: 10.1007/BF01227493. Errata: [160].

[162] I-Jeng Wang and JAMES C Spall. A constrained simultaneous perturbation stochastic approximation algorithm based on penalty functions. In *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, volume 1, pages 393–399. IEEE, 1999.

[163] Qi Wang and James C Spall. Discrete simultaneous perturbation stochastic approximation on loss function with noisy measurements. In *Proceedings of the 2011 American Control Conference*, pages 4520–4525. IEEE, 2011.

[164] Li Wei and Li Chunsheng. A novel system parameters design and performance analysis method for distributed satellite-borne sar system. *Advances in Space Research*, 50:272–281, 2012. doi: 10.1016/j.asr.2012.03.026.

[165] J.R. Wertz, D.F. Everett, and J.J. Puschell. *Space Mission Engineering: The New SMAD*, pages 298,299,316,486. Space technology library. Microcosm Press, Hawthorne, CA, first edition, 2011. ISBN 9781881883159.

[166] William R. Whittecar, Marc D. DiPrinzio, Lake A. Singh, and Matthew P. Ferringer. Petascale discovery of passively controlled satellite constellations for global coverage. pages 4013–4026. Univelt, 2015.

[167] William E Wiesel. *Spaceflight dynamics*. McGraw-Hill Science Engineering, 1997.

[168] William E. Wiesel. *Spaceflight Dynamics*. CreateSpace, Scotts Valley, CA, third edition, 2010. ISBN 978-1452879598.

[169] Yonas G. Woldesenbet, Gary G. Yen, and Biruk G. Tessema. Constraint handling in multiobjective evolutionary optimization. *IEE Transactions on Evolutionary Computation*, 13(3):514–525, 2009. doi: 0.1109/TEVC.2008.2009032.

[170] P. Xaypraseuth, R. Satish, and A. Chatterjee. Nisar spacecraft concept overview: Design challenges for a proposed flagship dual-frequency sar mission. In *2015 IEEE Aerospace Conference*, pages 1–11, March 2015. doi: 10.1109/AERO.2015.7118935.

[171] Dan Xue, Junfeng Li, Hexi Baoyin, and Fanghua Jiang. Reachable domain for spacecraft with a single impulse. *Journal of Guidance, Control, and Dynamics*, 33(3): 934–942, 2010. doi: 10.2514/1.43963.

[172] Gary G Yen and Zhenan He. Performance metric ensemble for multiobjective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 18(1): 131–144, 2014. doi: 10.1109/TEVC.2013.2240687.

[173] Ming F Yuen. Dilution of precision (dop) calculation for mission planning purposes. Master's thesis, NAVAL POSTGRADUATE SCHOOL MONTEREY CA, 2009.

[174] Costantinos Zagaris and Marcello Romano. Applied reachability analysis for spacecraft rendezvous and docking with a tumbling object. In *2018 Space Flight*

Mechanics Meeting, Kissimmee, FL, 2018. The American Institute of Aeronautics and Astronautics. doi: 10.2514/6.2018-2220.

[175] Manfred Zink, Markus Bachmann, Benjamin Brautigam, Thomas Fritz, Irena Hajnsek, Alberto Moreira, Birgit Wessel, and Gerhard Krieger. Tandem-x: the new global dem takes shape. *IEEE Geoscience and Remote Sensing Magazine*, 2(2):8–23, 2014.

Appendices

Appendix A

Heterogeneous launch manifestation

Launch manifestation for heterogeneous satellite constellations has been considered previously as part of the original DISCO methodology [28]. This work extends heterogeneous constellation manifestation to include the ability for satellites to maneuver to their final orbits. The propellant needed to perform the transfer, f_{ik}, is estimated using a two burn minimum energy Lambert targeter. The launch manifest problem can be modeled as a binary linear programming (BLP) problem and solved using a mixed integer linear programming solver such as CPLEX or Gurobi. The BLP has the following decision variables:

- x_{ij}: 1 if satellite i is assigned to launch vehicle j, 0 otherwise
- ω_{jk}: 1 if launch vehicle j is assigned to orbit k, 0 otherwise
- y_{jl}: 1 if launch vehicle j is assigned to launch location l, 0 otherwise

and constants:

- m_{jkl}: maximum allowable mass for launch vehicle j to orbit k from launch location l
- m_i: dry mass of satellite i, kg
- f_{ik}: propellant required to move satellite i from orbit k to its final orbit, capped at the maximum propellant mass for the satellite, kg

- r_{ik}: 1 if satellite i has sufficient propellant to get from orbit k to its final orbit, 0 otherwise
- α_{jkl}: 1 if launch vehicle j can go from launch location l to orbit k, 0 otherwise
- c_{jl}: cost in dollars to launch vehicle j from location l. Currently, cost is independent of location.
- I: total number of satellites
- J: total number of launch vehicles, equal to the number of launch vehicle types times I
- K: total number of unique orbits
- L: total number of launch locations

The following constraints must be satisfied by the launch manifest:

1. Each satellite is assigned to exactly one launch vehicle.

$$\sum_{j=1}^{J} x_{ij} = 1 \quad \forall i = 1, 2, \ldots, I \tag{A.1}$$

2. A launch vehicle is assigned to the same number of orbits as it is locations.

$$\sum_{l=1}^{L} y_{jl} = \sum_{k=1}^{K} \omega_{jk} \quad \forall j = 1, 2, \ldots, J \tag{A.2}$$

3. For a launch vehicle assigned to a specific launch location and orbit, the combined mass of all satellites assigned to the launch vehicle and the propellant required to get those satellites from the launch vehicle's orbit to their final orbit must be less than

or equal to the mass capacity of launch vehicle going to the assigned orbit from the assigned launch location.

A "big M" parameter M is defined such that if the launch vehicle is not assigned to either the orbit or the launch location, the constraint is satisfied automatically. Specifically, M is equal to the sum of the maximum wet mass of all of the satellites.

$$\sum_{i=1}^{I}(m_i + f_{ik})x_{ij} \leq m_{jkl} + M(1-\omega_{jk}) + M(1-y_{jl})$$
$$\forall j = 1,2,\ldots,J \quad \forall k = 1,2,\ldots,K$$
$$\forall l = 1,2,\ldots,L \tag{A.3}$$

4. If a launch vehicle is assigned to a specific launch location and orbit, launches from the specified launch location to the specified orbit using the specified launch vehicle must be allowed. Specifically, the launch vehicle must be available at the launch location and the inclination of the orbit must be in the allowable range for the launch location.

$$\omega_{jk} - \alpha_{jkl} + y_{jl} - 1 \leq 0$$
$$\forall j = 1,2,\ldots,J \quad \forall k = 1,2,\ldots,K \quad \forall l = 1,2,\ldots,L \tag{A.4}$$

5. If a satellite is assigned to a launch vehicle, and the launch vehicle is assigned to an orbit, then the satellite must be capable of reaching its final orbit from the launch orbit, subject to the constraints on the maximum propellant for the satellite.

$$\omega_{jk} - r_{ik} + x_{ij} - 1 \leq 0$$
$$\forall i = 1,2,\ldots,I \quad \forall j = 1,2,\ldots,J \quad \forall k = 1,2,\ldots,K \tag{A.5}$$

6. If at least one satellite is assigned to a launch vehicle, then that launch vehicle must be assigned to at least one orbit.

$$\sum_{i=1}^{I} x_{ij} - I \sum_{k=1}^{K} \omega_{jk} \leq 0 \quad \forall j = 1, 2, \ldots, J \tag{A.6}$$

7. A launch vehicle can be assigned to at most one orbit.

$$\sum_{k=1}^{K} \omega_{jk} \leq 1 \quad \forall j = 1, 2, \ldots, J \tag{A.7}$$

8. x_{ij} is binary.

$$x_{ij} \in \{0, 1\} \quad \forall i = 1, 2, \ldots, I \quad \forall j = 1, 2, \ldots, J \tag{A.8}$$

9. ω_{jk} is binary.

$$\omega_{jk} \in \{0, 1\} \quad \forall j = 1, 2, \ldots, J \quad \forall k = 1, 2, \ldots, K \tag{A.9}$$

10. y_{jl} is binary.

$$y_{jl} \in \{0, 1\} \quad \forall j = 1, 2, \ldots, J \quad \forall l = 1, 2, \ldots, L \tag{A.10}$$

The optimal launch manifest will minimize the cost of the launches assigned to launch locations, which is equivalent to minimizing the cost of all assigned launches. The objective function for the optimization is therefore:

$$\min \sum_{j=1}^{J} \sum_{l=1}^{L} c_{jl} y_{jl} \tag{A.11}$$

The BLP is solved to get the optimal launch manifest. When launch costs are calculated internally as part of a larger optimization problem, a time limit is imposed to prevent excessive computation time. If that time limit is reached, the program defaults to using

the cheapest single launch vehicle available for each satellite without combining multiple satellites onto a single launch vehicle.

Appendix B

Additional resilience formulations

This appendix develops mixed integer linear programming formulations for additional revisit metrics. All of the formulations require the binary satellite inclusion variable x_j and the binary access sufficiency variable Y_{ik} that were defined in subsection 5.2.2. Most formulations also use the constraint on the number of removals defined in Eq. (5.1).

B.1 Minimize/maximize average total access time

Total access time is one of the simplest metrics, as it merely requires summing the access times for each station to get the total access time per station, then averaging those values over all stations. The entire problem formulation is shown below. m sets the sign for the optimization, meaning $m = 1$ if minimization of the average total access time is desired and $m = -1$ if maximization of the average total access time is desired.

$$\text{Minimize} \sum_{k=1}^{n_g} \sum_{i=1}^{n_t-1} \frac{m Y_{ik} \Delta T_{ik}}{n_g}$$

such that:

$$x_j \in \{0,1\} \qquad \forall\ j = 1,2,\ldots,n_s$$

$$Y_{ik} \in \{0,1\} \qquad \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g$$

$$\sum_{j=1}^{n_s} x_j = n_s - n_r$$

$$Y_{ik} \leq \frac{\sum_{j=1}^{n_s} A_{ijk} x_j}{n_c} \qquad \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g$$

$$MY_{ik} \geq \sum_{j=1}^{n_s} A_{ijk} x_j - n_c + 1 \ \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g$$

This formulation has a total of $n_t \times n_g + n_s$ binary variables. Minimizing the average total access time is equivalent to maximizing the average total time with no access, and maximizing the average total access time is equivalent to minimizing the average total time with no access. For the purposes of finding the worst case scenario, minimizing the average total access time is desired.

B.2 Minimize/maximize average number of gaps and passes

Both the number of gaps in coverage and the total time a station is without access are a function of the included assets. Therefore, the average gap length for a particular ground station must be nonlinear, since it requires dividing the total gap time by the number of gaps. Instead, a means of calculating the number of gaps is presented. If desired, this metric can be used in conjunction with the total access time to approximate the average revisit time, though performance will be subject to the weighting of the two metrics.

To find the number of gaps (or passes) for each station, the number of times a station switches from having access to having no access (or vice versa) must be found. A pair of binary variables, D_{ik}^{+} and D_{ik}^{-} track these changes in state. D_{ik}^{+} is one if the station k goes from no access at time T_{ik} to access at time $T_{(i+1)k}$. D_{ik}^{-} is one if the station k goes from access at time T_{ik} to no access at time $T_{(i+1)k}$. Both variables are zero when there is not a

change in access. These relationships are guaranteed by the following constraints:

$$D_{ik}^{+} - D_{ik}^{-} = Y_{(i+1)k} - Y_{ik} \forall i = 1, 2, \ldots, n_t - 1, k = 1, 2, \ldots, n_g \tag{B.1}$$

$$D_{ik}^{+} + D_{ik}^{-} \leq 1 \forall i = 1, 2, \ldots, n_t - 1, k = 1, 2, \ldots, n_g \tag{B.2}$$

Eq. (B.1) requires that $D_{ik}^{+} = 1$ and $D_{ik}^{-} = 0$ when coverage begins, and $D_{ik}^{+} = 0$ and $D_{ik}^{-} = 1$ when coverage ends. By itself, it requires only that $D_{ik}^{+} = D_{ik}^{-}$ when there is no change in access. Since Eq. (B.2) prevents both variables from being one at the same time step, both must be equal to zero when there is no change in access.

With these constraints in place, the optimization problem for the gap counting case can be formulated. The number of gaps for a ground station is equal to the number of switches to a period of no access, plus one if the scenario starts in a period of no access.

$$\text{Minimize } \frac{m}{n_g} \sum_{k=1}^{n_g} \left((1 - Y_{1k}) + \sum_{i=1}^{n_t - 1} D_{ik}^{-} \right)$$

such that:

$$x_j \in \{0, 1\} \qquad \forall\ j = 1, 2, \ldots, n_s$$

$$Y_{ik} \in \{0, 1\} \qquad \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g$$

$$D_{ik}^{+} \in \{0, 1\} \qquad \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g$$

$$D_{ik}^{-} \in \{0, 1\} \qquad \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g$$

$$\sum_{j=1}^{n_s} x_j = n_s - n_r$$

$$Y_{ik} \leq \frac{\sum_{j=1}^{n_s} A_{ijk} x_j}{n_c} \qquad \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g$$

$$M Y_{ik} \geq \sum_{j=1}^{n_s} A_{ijk} x_j - n_c + 1 \qquad \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g$$

$$D_{ik}^{+} - D_{ik}^{-} = Y_{(i+1)k} - Y_{ik} \qquad \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g$$

$$D_{ik}^{+} + D_{ik}^{-} \leq 1 \qquad \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g$$

The formulation above counts the number of gaps, but the number of passes could be used instead by replacing the objective function with

$$\text{Minimize } \frac{m}{n_g} \sum_{k=1}^{n_g} \left(Y_{1k} + \sum_{i=1}^{n_t - 1} D_{ik}^{+} \right) \tag{B.3}$$

The choice of metric that yields the worst case scenario is not obvious when using number of gaps and passes. In some cases, removing some satellites can break up an existing pass, increasing both the number of gaps and number of passes. In other cases, removing satellites can remove a pass completely, decreasing the number of passes and gaps. The latter case is more likely to occur than the former when dealing with small constellations with discontinuous coverage, as a removal is likely to completely remove a pass rather than simply breaking it apart. In cases where the nominal coverage is continuous or nearly continuous, the number of passes and gaps will increase as coverage degrades. This formulation has a total of $(3n_t - 2) \times n_g + n_s$ binary variables.

B.3 Minimize maximum revisit times

The maximum revisit time of a ground station is the longest period for which that station is without coverage. The overall maximum revisit times is the largest gap in coverage for any station in the scenario. The minimization formulation differs from the maximization formulation, though the accumulator variable has the same definition and constraints as described in subsection 5.2.2.

The length of the largest gap is equal to the largest value of a_{ik}. To find this value, the variable $a_{max} \in \mathbb{R}_{\geq 0}$ is used. a_{max} must be greater than all values of a. Since a_{max} is minimized, no further constraints are needed as a_{max} will be driven to the largest value of a and will reduce that value as much as possible. The total formulation is

Minimize a_{max}

such that:

$$
\begin{aligned}
x_j &\in \{0,1\} && \forall\ j = 1,2,\ldots,n_s \\
Y_{ik} &\in \{0,1\} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
a_{ik} &\in \mathbb{R}_{\geq 0} && \forall\ i = 1,2,\ldots,n_t - 1,\ k = 1,2,\ldots,n_g \\
a_{max} &\in \mathbb{R}_{\geq 0} \\
\sum_{j=1}^{n_s} x_j &= n_s - n_r \\
Y_{ik} &\leq \frac{\sum_{j=1}^{n_s} A_{ijk} x_j}{n_c} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
MY_{ik} &\geq \sum_{j=1}^{n_s} A_{ijk} x_j - n_c + 1 && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
a_{1k} &\geq \Delta T_{1k} - M_2 Y_{1k} && \forall\ k = 1,2,\ldots,n_g \\
a_{1k} &\leq \Delta T_{1k} + M_2 Y1k && \forall\ k = 1,2,\ldots,n_g \\
a_{ik} &\geq a_{(i-1)k} + \Delta T_{ik} - M_2 Y_{ik} && \forall\ i = 2,3,\ldots,n_t - 1,\ k = 1,2,\ldots,n_g \\
a_{ik} &\leq a_{(i-1)k} + \Delta T_{ik} + M_2 Y_{ik} && \forall\ i = 2,3,\ldots,n_t - 1,\ k = 1,2,\ldots,n_g \\
a_{ik} &\leq M_2(1 - Y_{ik}) && \forall\ i = 1,2,\ldots,n_t - 1,\ k = 1,2,\ldots,n_g \\
a_{max} &\geq a_{ik} && \forall\ i = 1,2,\ldots,n_t - 1,\ k = 1,2,\ldots,n_g
\end{aligned}
$$

This formulation has a total of $n_t \times n_g + n_s$ binary variables and $(n_t - 1) \times n_g + 1$ continuous variables.

B.4 Minimize/maximize average of maximum revisit times

Finding the average of the maximum revisit times is a similar process to finding the maximum. The development of the accumulator variable a is the same as in the maximum case. However, each ground station has a separate maximum variable $a_{max,k}$ in the average case. Again, the formulation depends on whether the problem is a maximization or a minimization problem.

Minimization: The maximum value for each station must be greater than or equal to each a value for that station. The objective function is the average of the maximum values. The entire problem formulation is

$$\text{Minimize } \frac{1}{n_g}\sum_{k=1}^{n_g} a_{max,k}$$

such that:

$$\begin{aligned}
x_j &\in \{0,1\} && \forall\ j = 1,2,\ldots,n_s \\
Y_{ik} &\in \{0,1\} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
a_{ik} &\in \mathbb{R}_{\geq 0} && \forall\ i = 1,2,\ldots,n_t-1,\ k = 1,2,\ldots,n_g \\
a_{max,k} &\in \mathbb{R}_{\geq 0} && \forall\ k = 1,2,\ldots,n_g \\
\sum_{j=1}^{n_s} x_j &= n_s - n_r \\
Y_{ik} &\leq \frac{\sum_{j=1}^{n_s} A_{ijk}x_j}{n_c} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
MY_{ik} &\geq \sum_{j=1}^{n_s} A_{ijk}x_j - n_c + 1 && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
a_{1k} &\geq \Delta T_{1k} - M_2 Y_{1k} && \forall\ k = 1,2,\ldots,n_g
\end{aligned}$$

$$
\begin{aligned}
a_{1k} &\leq \Delta T_{1k} + M_2 Y 1k && \forall\ k = 1, 2, \ldots, n_g \\
a_{ik} &\geq a_{(i-1)k} + \Delta T_{ik} - M_2 Y_{ik} \forall\ i = 2, 3, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{ik} &\leq a_{(i-1)k} + \Delta T_{ik} + M_2 Y_{ik} \forall\ i = 2, 3, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{ik} &\leq M_2 (1 - Y_{ik}) && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{max,k} &\geq a_{ik} && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g
\end{aligned}
$$

This formulation has a total of $n_t \times n_g + n_s$ binary variables and $n_t \times n_g$ continuous variables.

Maximization: As with the maximum revisit time case, a binary decision variables δ_{ik} is used. For each value of $a_{max,k}$, a single constraint that forces $a_{max,k}$ to be less than or equal to one of the values of a_{ik} is enforced. The resulting constraints are

$$a_{max,k} \leq a_{ik} + (1 - \delta_{ik}) M_2\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \tag{B.4}$$

$$\sum_{i=1}^{n_t - 1} \delta_{ik} = 1 \forall\ k = 1, 2, \ldots, n_g \tag{B.5}$$

The maximum values are then averaged to get the average of the maximum revisit times. The full formulation is

$$\text{Minimize } \frac{-1}{n_g} \sum_{k=1}^{n_g} a_{max,k}$$

such that:

$$
\begin{aligned}
x_j &\in \{0, 1\} && \forall\ j = 1, 2, \ldots, n_s \\
Y_{ik} &\in \{0, 1\} && \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g \\
a_{ik} &\in \mathbb{R}_{\geq 0} && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g
\end{aligned}
$$

$$
\begin{aligned}
a_{max,k} &\in \mathbb{R}_{\geq 0} && \forall\ k = 1, 2, \ldots, n_g \\
\delta_{ik} &\in \{0, 1\} && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
\sum_{j=1}^{n_s} x_j &= n_s - n_r \\
Y_{ik} &\leq \frac{\sum_{j=1}^{n_s} A_{ijk} x_j}{n_c} && \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g \\
MY_{ik} &\geq \sum_{j=1}^{n_s} A_{ijk} x_j - n_c + 1 && \forall\ i = 1, 2, \ldots, n_t,\ k = 1, 2, \ldots, n_g \\
a_{1k} &\geq \Delta T_{1k} - M_2 Y_{1k} && \forall\ k = 1, 2, \ldots, n_g \\
a_{1k} &\leq \Delta T_{1k} + M_2 Y1k && \forall\ k = 1, 2, \ldots, n_g \\
a_{ik} &\geq a_{(i-1)k} + \Delta T_{ik} - M_2 Y_{ik} \forall\ i = 2, 3, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{ik} &\leq a_{(i-1)k} + \Delta T_{ik} + M_2 Y_{ik} \forall\ i = 2, 3, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{ik} &\leq M_2(1 - Y_{ik}) && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
a_{max,k} &\leq a_{ik} + (1 - \delta_{ik}) M_2 && \forall\ i = 1, 2, \ldots, n_t - 1,\ k = 1, 2, \ldots, n_g \\
\sum_{i=1}^{n_t - 1} \delta_{ik} &= 1 && \forall\ k = 1, 2, \ldots, n_g
\end{aligned}
$$

This formulation has a total of $(2n_t - 1) \times n_g + n_s$ binary variables and $n_t \times n_g$ continuous variables.

B.5 Minimize/maximize number of removals to reach threshold

Previous formulations focus on finding the best or worst performance achievable for a fixed number of satellite removals. Alternatively, it is possible to minimize or maximize the number of assets that must be removed before a performance threshold is violated. For example,

rather than maximizing the maximum revisit time for a constellation of satellites given four removals, a formulation could minimize the number of satellites such that the maximum revisit time is more than 25 minutes. The formulation will differ slightly for each metric. The formulation for finding the minimum number of removals such that the maximum revisit time exceeds some threshold is given here.

The formulation is as given in subsection 5.2.2 except for a change in the objective function, a change in Eqs. (5.10) and (5.11), and the removal of the constraint described by Eq. (5.1). Because the objective is to minimize the number of removals, the objective function is

$$\text{Minimize} \sum_{j=1}^{n_s} -x_j \tag{B.6}$$

The largest value in a is the maximum revisit time. If any value of a_{ik} exceeds the threshold h, the requirement is satisfied. Since the threshold only needs to be exceeded by a single value of a_{ik}, the constraint is

$$a_{ik} + (1 - \delta_{ik})h \geq h \forall i = 1, 2, \ldots, n_t - 1, \;\; k = 1, 2, \ldots, n_g \tag{B.7}$$

Because the threshold can be exceeded by more than one value of a_{ik}, Eq. (5.11) can be relaxed to an inequality constraint if desired. If at least one value of a_{ik} exceeds the threshold for some combination of satellite removals, that set of removals is valid. The full formulation is

$$\text{Minimize} \sum_{j=1}^{n_s} -x_j$$

such that:

$$x_j \in \{0, 1\} \qquad \forall \; j = 1, 2, \ldots, n_s$$

$$
\begin{aligned}
Y_{ik} &\in \{0,1\} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
a_{ik} &\in \mathbb{R}_{\geq 0} && \forall\ i = 1,2,\ldots,n_t-1,\ k = 1,2,\ldots,n_g \\
\delta_{ik} &\in \{0,1\} && \forall\ i = 1,2,\ldots,n_t-1,\ k = 1,2,\ldots,n_g \\
Y_{ik} &\leq \frac{\sum_{j=1}^{n_s} A_{ijk} x_j}{n_c} && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
MY_{ik} &\geq \sum_{j=1}^{n_s} A_{ijk} x_j - n_c + 1 && \forall\ i = 1,2,\ldots,n_t,\ k = 1,2,\ldots,n_g \\
a_{1k} &\geq \Delta T_{1k} - M_2 Y_{1k} && \forall\ k = 1,2,\ldots,n_g \\
a_{1k} &\leq \Delta T_{1k} + M_2 Y_{1k} && \forall\ k = 1,2,\ldots,n_g \\
a_{ik} &\geq a_{(i-1)k} + \Delta T_{ik} - M_2 Y_{ik} && \forall i = 2,3,\ldots,n_t-1,\ k = 1,2,\ldots,n_g \\
a_{ik} &\leq a_{(i-1)k} + \Delta T_{ik} + M_2 Y_{ik} && \forall i = 2,3,\ldots,n_t-1,\ k = 1,2,\ldots,n_g \\
a_{ik} &\leq M_2(1 - Y_{ik}) && \forall i = 1,2,\ldots,n_t-1,\ k = 1,2,\ldots,n_g \\
h &\leq a_{ik} + (1 - \delta_{ik})h && \forall i = 1,2,\ldots,n_t-1,\ k = 1,2,\ldots,n_g \\
\sum_{i=1}^{n_t-1} \sum_{k=1}^{n_g} \delta_{ik} &\geq 1
\end{aligned}
$$

This formulation has a total of $(2n_t-1)\times n_g+n_s$ binary variables and $(n_t-1)\times n_g$ continuous variables.

B.6 Probability inclusion

Because the resilience formulations are designed to find the best or worst case scenarios, there is no consideration given to the likelihood that a particular combination of losses will occur. However, it may be desirable to find the best or worst case scenario that has some likelihood of occurrence rather than a fixed number of losses. Each satellite has some probability of failure $p_j \in (0,1)$. If the set S_+ defines the set of satellites that have not been removed and

S_- defines the set of satellites that have been removed, the probability of occurrence p_e for the exact combination of removals is

$$p_e = \prod_{q \in S_-} p_q \prod_{r \in S_+} (1 - p_r) \tag{B.8}$$

p_e is called the exact combination probability. Similarly, the probability of occurrence p_a that at least the satellites in S_- are removed is

$$p_a = \prod_{q \in S_-} p_q \tag{B.9}$$

p_a is called the threshold combination probability. Because products are nonlinear, Eqs. (B.8) and (B.9) must be converted to log space:

$$\log(p_e) = \sum_{q \in S_-} \log(p_q) + \sum_{r \in S_+} \log(1 - p_r) \tag{B.10}$$

$$\log(p_a) = \sum_{q \in S_-} \log(p_q) \tag{B.11}$$

A satellite is in S_- if $x_j = 0$ and in S_+ is $x_j = 1$. If the probability threshold for the analysis is p_t, it is required that $\log(p_e) \geq \log(p_t)$ if the exact combination probability is used. It is required that $\log(p_a) \geq \log(p_t)$ if the threshold combination probability is used. These constraints can be written as

$$\log(p_e) = \sum_{j=1}^{n_s} (1 - x_j) \log(p_j) + x_j \log(1 - p_j) \geq \log(p_t) \tag{B.12}$$

$$\log(p_a) = \sum_{j=1}^{n_s} (1 - x_j) \log(p_j) \geq \log(p_t) \tag{B.13}$$

Either Eq. (B.12) or Eq. (B.13) can be used to replace Eq. (5.1) in any resilience analysis. Since $p_a \geq p_e$, Eq. (B.13) is less restrictive.