Multiobjective Optimization of a Radio Telescope Array with Site Constraints

by

Babak Cohanim

Submitted to the Department of Aeronautics/Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics/Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2004

Author .....
Department of Aeronautics/Astronautics
February 1, 2004

Certified by .....
Olivier de Weck
Assistant Professor of Aero/Astro & Engineering Systems
Thesis Supervisor

Certified by .....
Jacqueline Hewitt
Professor of Physics, Director of Center for Space Research
Thesis Supervisor

Accepted by .....
Edward M. Greitzer
Chairman, Committee on Graduate Students

AERO

Multiobjective Optimization of a Radio Telescope Array with Site Constraints

by

Babak Cohanim

Submitted to the Department of Aeronautics/Astronautics
on February 1, 2004, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics/Astronautics

Abstract

The next generation of radio telescope interferometric arrays requires careful design of the array configuration to optimize the performance and minimize the cost of the overall system while adhering to site constraints. A framework, based on genetic algorithms, for rapid exploration and optimization of the objective space pertaining to multiple objectives has been developed. A large space of possible designs has been evaluated for 27-, 60-, 100-, and 160-station arrays. The 27-station optimizations can be compared to the well-known VLA case, and the larger array designs apply to arrays currently under design such as LOFAR, ATA, and the SKA.

In the initial implementation of the framework, designs were evaluated with respect to two metrics, array imaging performance and the length of cable necessary to connect the stations which adhere to LOFAR design objectives and parameters. Imaging performance is measured by the degree to which the sampling of the uv plane is uniform. The length of cable used to connect stations was determined to be a key cost parameter in the construction of the array. Site constraints are a key issue in any ground-based system and have been included to ascertain the effects on the objective space. Results will show non-dominated configurations for the metrics chosen as one metric is emphasized over another, as well as a comparison of how site constraints affect the overall design and convergence of the genetic algorithm simulations.

The framework is general, and may be applied to other design goals and issues, such as particular schemes for sampling the uv plane, array robustness, and phased deployment of arrays.

Thesis Supervisor: Olivier de Weck
Title: Assistant Professor of Aero/Astro & Engineering Systems

Thesis Supervisor: Jacqueline Hewitt
Title: Professor of Physics, Director of Center for Space Research

Acknowledgments

First, I would like to thank my advisor Jacqueline Hewitt for giving me the opportunity to be able to do research in both the fields of Astronomy and Engineering. It was my desire to be able to work in and learn about research methods of both a scientist and an engineer. Jackie has given me valuable insight into what it takes to be a scientist and how to interact with other scientists. For all the opportunities and knowledge she has given me, I am thankful.

Second, I would like to thank my advisor Olivier de Weck for the many hours of discussion on my research and thesis. The knowledge I received of multiobjective optimization and complex systems was invaluable in conducting my thesis research.

I would like to thank Dr. Wesley Harris for getting me started at MIT. The caring and mentorship he showed to me initially helped me to get my foot in the door and take action about my own path at MIT.

I would like to thank Miguel Morales, Divya Oberoi, Colin Lonsdale, Sheperd Doeleman, Roger Cappallo, Ramesh Bhat, and the other research scientists at MIT Haystack Observatory for all the help they provided while I was conducting my research. Their help and patience in discussions and correspondence helped me to learn about LOFAR and the intricacies of the problem I was attempting to solve for my thesis.

I would like to thank my officemates Miriam Krauss, Judd Bowman, Nick Fotopoulos, and Clayton Featherstone for helping to keep my sanity while writing my thesis and doing classwork. I would also like to acknowledge all the faculty, postdoctoral fellows, and graduate students in the Center for Space Research and Space Systems Laboratory that I worked and interacted with throughout my time as a Master's student at MIT.

Finally I would like to thank my family and friends for all their support and advice, and shaping me into the person I am today. Without them none of this would have been possible.

Contents

List of Tables

List of Figures

Nomenclature

D_{NU}	Minimum distance to nadir-utopia point
J	Design objective
L_{cable}	Cable length, km
M	Array performance metric, %
$N_{stations}$	Number of stations
N_{uv}	Total number of uv points
$N_{uvactual}$	Filled uv points
badzone	Area in which an array element cannot be placed
erate	Elitism rate, %
gridstep	Site constraint grid dimension, km
mrate	Mutation rate, %
ngen	Total number of generations
percent	Percentage of coverage in site constraints, %
popsize	Population size used in optimization
uv	Interferometric baseline vectors, kλ
xrate	Crossover rate, %
xy	Ground positions of stations, km
xygrid	Maximum radius in xy plane, km
λ	Wavelength, m

Chapter 1

Introduction

1.1 Science Motivation

Probing deeper back into time and looking for fainter objects against a very bright background present challenges for the next generation of radio telescopes. Large single dish apertures, like the Arecibo Observatory [1] obtain high resolution from having a very large dish size (Arecibo has the largest dish in the world at 305 meters). Equation 1.1 shows the relationship between beamwidth (angular resolution) and dish diameter as a function of observing wavelength.

$$\theta \approx \frac{\lambda}{D} \tag{1.1}$$

where θ is the minimum resolvable angle (resolution), λ the observing wavelength, and D is the aperture diameter. For Arecibo, the highest angular resolution is achieved at its highest frequency (10 GHz) and is near 20 arcseconds. It can easily be seen that achieving higher resolutions would require a very large dish.

Interferometers are many dishes working together to act as one large telescope [21]. The Very Large Array [20] uses 27 dishes working together with a longest baseline of 36 kilometers, which gives a resolution of 0.05 arcseconds at its highest operating frequency of 40 GHz. As can be seen, the increase in baselines offered by interferometry makes it possible to probe at much finer resolutions without building infeasible

large single dish telescopes for the job.

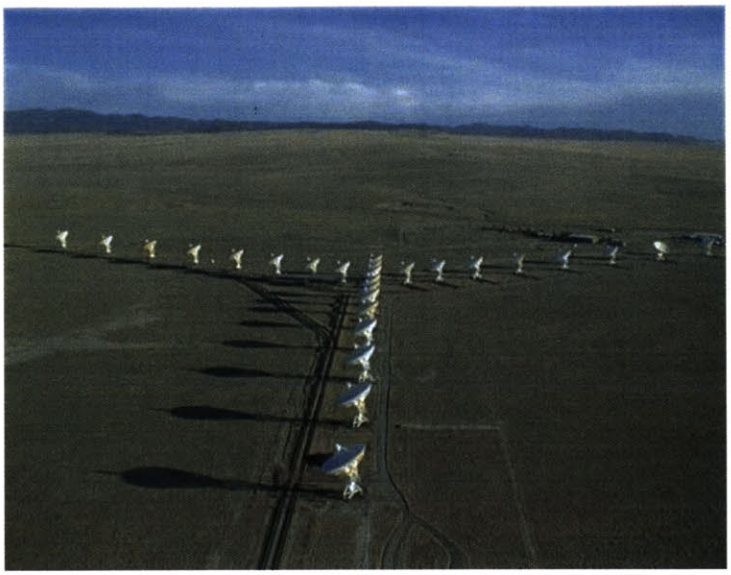


Figure 1-1: Very Large Array

The next generation of radio telescope interferometric arrays will require careful planning of the array configuration to optimize the performance and reduce the cost of the overall system. The placement of the antennas determines the sampling of the Fourier transform of the sky brightness distribution and hence the fidelity of the image computed from the interferometric data [20]. One dimensional arrays use earth rotation aperture synthesis to sample the uv plane over time as the earth rotates, and for these arrays minimally redundant antenna configurations are relatively easily found. However, for arrays which are required to operate over a wide range of declinations and for arrays where instantaneous capabilities are important, the much more difficult problem of two-dimensional configurations of antennas must be considered. The point spread function (PSF), or beam, of an array is the Fourier transform of the cross-correlation function (uv plane) of the station placements (xy plane). An example is shown in Figure 1-2 using a VLA-like design. The beam is easily computed from the coordinates of the antennas. Unfortunately, there is no analytic solution to the inverse problem of creating an array configuration for the desired beam. Both the cross correlation function and the gridded Fast Fourier Transform lose informa-

tion when computed. Inverting the process would require a deconvolution in both instances, making it highly improbable to recover the station placements in the xy plane [18].

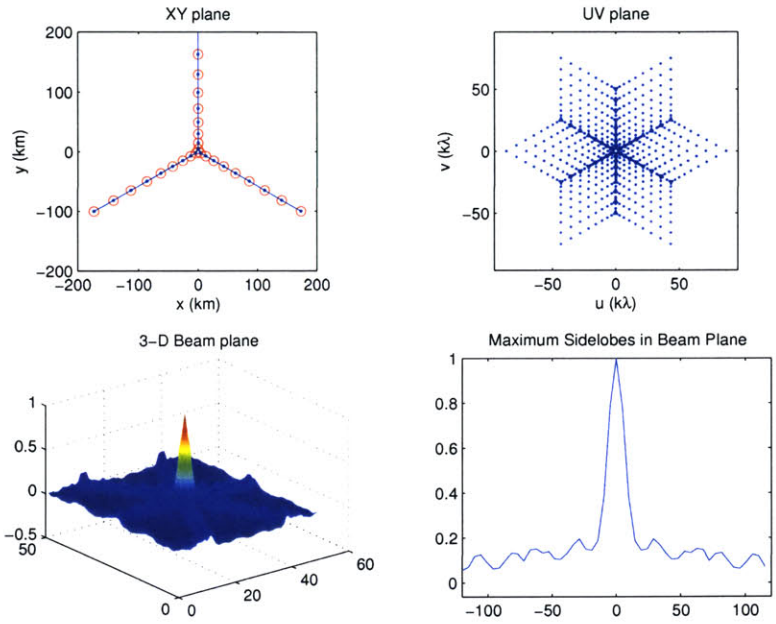


Figure 1-2: Sample xy to uv to PSF. A sample 27-station VLA configuration (upper-left) with its corresponding uv distribution (upper-right) and point spread function beam shape (lower-left) and maximum normalized beam sidelobes (lower-right).

Performance may not be the only design objective of an interferometric array. Cost is often an issue. The placement of the antennas also affects the cost of the array by determining the costs of power and signal distribution, site preparation, roads, and other infrastructure items. This becomes all the more complicated in a non-ideal world where the array layout is subject to its natural and human surroundings. Mountains, rivers, and cities are in the scope of the development area of proposed next generation telescopes. Adapting to and avoiding these environments will be necessary in the construction of these large telescopes. All of these considerations lead to trade-offs that must be considered in the design.

Thus a new framework is needed to optimize upon these multiple objectives, taking into account the constraints, and aid in the decision making process of how the array should be laid out and what infrastructure may be required.

1.2 Array Design Optimization Challenge

A method is developed in this thesis for rapid exploration of the objective space pertaining to multiple objectives. Previous work by Cornwell [5] suggests a highly nonlinear objective space with respect to obtaining a beam pattern from a given array configuration (refer back to Section 1.1). Gradient search techniques are prone to getting trapped in local minima. Thus, as has been shown by Boone [3], Cornwell [5], Keto [11], and several others, the use of heuristic techniques such as simulated annealing, neural networks, and genetic algorithms, which are all better at handling nonlinear objective spaces, need to be used, but also require greater computational resources. Genetic algorithms [10] produce many configurations throughout a single simulation run, giving a Pareto front [24] of optimal solutions at once. Simulated annealing and neural networks converge to a single solution, requiring several simulation runs of changing the weighting functions to determine a Pareto front. The method presented here will aid array designers in choosing an optimal configuration for their array depending on objectives that they can choose and trade upon.

1.2.1 Ground-Based Arrays

The first large two-dimensional radio astronomical array was the Very Large Array [20] which has a three-armed Y-shaped configuration. This configuration was first considered because it incorporated straight lines of antennas yet also distributed the antennas over a two dimensional region. The ability of such a configuration to cover the uv plane was supported by empirical studies of the transfer function, and positions within the Y were chosen through an optimization procedure. In more recent work various procedures for optimizing the performance of two-dimensional arrays have been developed by Boone [3], Cornwell [5], Keto [11]. These studies have focused on array performance as the sole objective. Future radio interferometric arrays now being designed, such as the Allen Telescope Array [2], the Square Kilometer array [16], and Low Frequency Array [14], take advantage of advances in signal processing to construct a large aperture from a large number of relatively small antenna

elements, or stations. For these arrays, the cost of connecting the stations can be a significant fraction of the total cost. As part of the design effort for LOFAR, an optimization has been done on two-dimensional configurations with two objectives considered simultaneously, array performance and cable length.

LOFAR

The Low Frequency Array [14] is being designed to probe the frequency range between 10 MHz and 240 MHz. LOFAR is to be built from many small dipole antennas which are bunched into clusters (containing 8-9 elements) and stations (containing 14-15 clusters). There will be approximately 100 stations in LOFAR, narrowed down from a range of 60 to 160 stations. LOFAR will also have large baselines, with stations being contained inside a 200 kilometer radius. The study in this thesis is based upon station placement optimization to increase performance and minimize cable laying costs associated with developing arrays on the same scale as LOFAR, such as the Square Kilometer Array, or SKA [16].

The data transfer system for LOFAR will be made of fiber optics buried beneath the ground which carry the interferometric data back to a central processing facility. Cable laying is seen as a major cost in the construction of the array, thus minimizing cable length is an important design objective. A wireless communication network would be difficult to use for data return because of the high bandwidth required.

LOFAR based objectives will be used in this thesis as a case study for the designed framework. It should be noted that the objectives that are considered are specific for LOFAR and the results reflect this. Array designs produced using different objectives will produce different configurations than presented here.

1.2.2 Space-Based Arrays

Space-based arrays compare and contrast to ground-based arrays in a number of ways. An obvious difference is that ground-based arrays are located within the earth's atmosphere, while space-based arrays are in a vacuum. The earth's atmosphere attenuates

many frequencies in the electromagnetic spectrum, and being able to bypass this is a great advantage to space-based arrays.

The uv plane of arrays is determined by the cross correlations of the elements with one another. Being able to reconfigure arrays allows them to alter there uv planes, which can be advantageous when using interferometric telescopes. Reconfiguration for space-based arrays is handled differently than for ground-based arrays. If a ground-based array is reconfigurable, such as the VLA [20], it is usually done by moving the elements of the arrays along fixed paths (for the VLA this is railroad tracks). In space, arrays fly in orbits, be it around the earth, sun, or other interstellar objects. Reconfiguring free flying arrays can be done by either using thrusters, or clever orbital dynamics which constantly reconfigures the array based on the slightly different orbits that the array elements take. Golay [9] arrays with threefold symmetry are interesting configurations to consider for space-based arrays because the threefold symmetry allows for orbits which will keep a stable configuration with fuel only needed for orbit corrections occurring from perturbation effects.

Filling the uv plane is important to both ground-based arrays and space-based arrays. On the ground this is done by earth-synthesis rotation [21]. In space this is done by tracking during an orbit or rotating the array. Integration times of viewing a single object are different depending on location as well. On the ground, tracking can be done from horizon to horizon at best. For space-based arrays orbiting the earth, integration times may only last as long as the orbital period of the array around the earth if the earth comes between the array and the source. In sun-orbiting arrays, the array elements are usually pointed away from the sun because it is such a strong source across the entire spectrum. Integration times for these arrays may be days to weeks depending on how far out the orbits are located.

The cost of space-based arrays is definitely the reason that there are so few of them compared to ground-based arrays. The cost of a launch is very high. Space-based arrays need to work correctly the first time, because unlike ground-based arrays, they do not get serviced. This means that a lot of testing and space hardening is required, which takes a lot of time, manpower, and money.

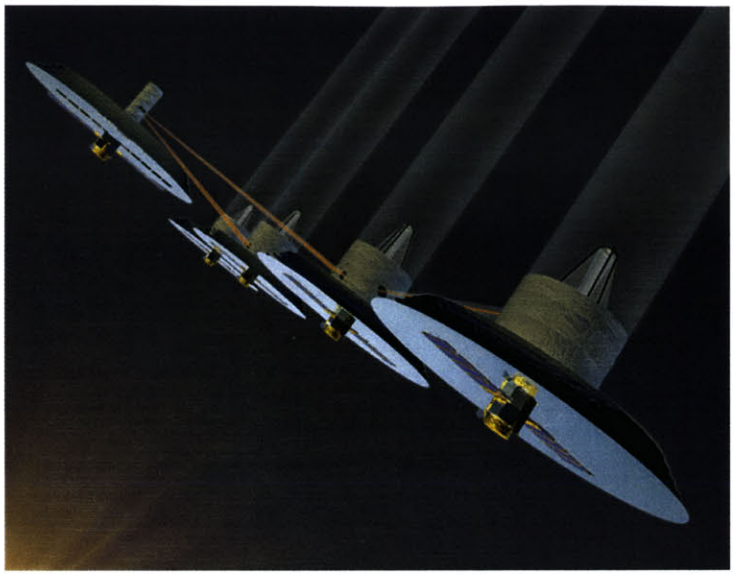


Figure 1-3: Terrestrial Planet Finder is an example of a space-based optical interferometer.

Though the models developed as an example for this study are based on ground arrays, it is possible to use this method for designing optimal array configurations for space-based arrays with the development of appropriate models. This will be an interesting application for future work.

1.3 Thesis Objectives and Outline

The goal of this thesis is to develop and validate a general framework for multiobjective optimization of radio telescope array configurations subject to site constraints.

Though the framework is general for any interferometric array, it is necessary to demonstrate the capabilities of the framework through a case study. For this purpose objectives have been chosen which are based upon the design efforts for LOFAR [14]. Here the key issues are uv visibility coverage as a performance metric and station connecting cable length as a predictor of significant cost in the design. This section will discuss the choice of metrics, the mathematical problem formulation, and the overall outline of this thesis.

1.3.1 Potential Objectives for Radio Telescope Arrays

Radio telescope arrays have a large number of objectives upon which they can be designed. For almost every telescope there needs to be a definition of performance. What this definition is can vary greatly depending on the science. Boone [3] [4] optimized the beam shape. Cornwell [5] and Keto [11] optimized filling the uv plane as completely as possible. Science that tries to detect very faint or distant objects may need to optimize on signal to noise, sensitivity, high dynamic range, and low RMS sidelobes. The choice of exactly which metric to use depends on the array designer. There are possibly metrics that can be chosen which have not been discovered yet.

Non-performance issues are also a very important part of the design of any complex system, especially radio telescope arrays. Cost is a major issue in any project. Objectifying cost may be easy or difficult depending on the complexity of the system. There are times when the major cost driver is already known, such as the cable laying costs associated with LOFAR [14]. The number of elements to be used in an array or the size of the elements may be a good objective to optimize cost upon. Difficulty comes when the cost is dependent on several things or is difficult to objectify because there is no good basis of knowledge for the complexity of the system. Other non-performance issues such as array robustness or extensibility may be important. As long as an issue can be quantified it can be used as a metric to be considered in a multiobjective framework.

Different scientists and designers can have many ideas for arrays, making the list of objectives to choose from seem endless. Choosing the objectives to optimize is difficult in itself as there are many of them to choose from. Figuring out which ones are important for different science objectives is an interesting topic and will be discussed in the section on Future Work (Section 6.3). The choices used in this thesis are based on LOFAR [14] and are given below.

LOFAR: Maximizing UV Density/Minimizing Cable Length

The LOFAR design is well suited for a case study because it offers two concrete opposing objectives to be traded off, array performance and array cost. There are many metrics for array performance. Work done by Boone [4], Cornwell [5], Kogan [12], and Woody [22], shows that minimizing the sidelobe levels of the point spread function for an instantaneous, monochromatic observation centered at zenith allows for better imaging quality. This is what has been done here; details are included in Section 2.4.1. LOFAR is to be a sparse array spanning many kilometers. Minimizing cable laying costs is directly related to minimizing cable length. As site information matures, the actual cost of laying cable through a specific region can be included, but for the purposes of this study, minimizing the cable length is a good corollary.

1.3.2 Site Constraints

An interesting consideration in the design of ground-based arrays is the land that they are built on. For the next generation of arrays the land that will be spanned will be hundreds of kilometers across. It is difficult to find a stretch of Earth that big that has perfectly flat, unpopulated land. The idea of site constraints in optimization of arrays is not new. Boone [3] used site constraints but dealt with them slightly differently than what will be done in this study. Chapters 4 & 5 will show results of simulations done using site constraints. Section 2.4.3 will describe how site constraints were handled in this study.

Topography is also an important issue. Elevation changes, radio frequency interference zones, even the earth's curvature make site constraints a difficult problem. For this thesis the issue of site constraints has been simplified to just binary 'goodzones' and 'badzones' where elements can be placed. Further discussion of this is in Section 2.4.3.

1.3.3 Mathematical Problem Formulation

A formal mathematical formulation of the framework is given in Equation 1.2. All metrics are defined as a minimization (doing a maximization is done by changing the sign in the minimization problem):

$$\min \left\{ J_i \right\} \tag{1.2}$$

where J is a design objective

$$\begin{aligned} J_i &= f(x_i, y_i, N_{stations}, \lambda, \ldots) \\ s.t.\ \sqrt{x_i^2 + y_i^2} &\leq \max(site\ radius) \\ x_i, y_i &\notin badzones \\ i &= 1, 2, \ldots, N_{stations} \end{aligned}$$

A formal mathematical formulation of the specific problem is given in Equation 1.3. Both metrics are to be minimized. Further discussion of the mathematics behind the objectives themselves will be presented in Sections 2.4.1 and 2.4.2. M is the array performance metric while L_{cable} is the cable length. Both are a function of the xy positions of the stations, the total number of stations, and the performance metric is also a function of the operating frequency. All simulations for this study were done with the stations within a radius of 200 km in accordance with the LOFAR [14] baseline design.

$$\min \left\{ \begin{array}{c} M \\ L_{cable} \end{array} \right\} \tag{1.3}$$

where

$$M = f(x_i, y_i, N_{stations}, \lambda)$$

$$L_{cable} = f(x_i, y_i, N_{stations})$$

$$i = 1, 2, \ldots, N_{stations}$$

$$s.t. \sqrt{x_i^2 + y_i^2} \leq 200\ km$$
$$x_i, y_i \notin badzones$$

1.3.4 Thesis Overview

Figure 1-4 shows a road map for this thesis. Given below is further detail on the chapters.

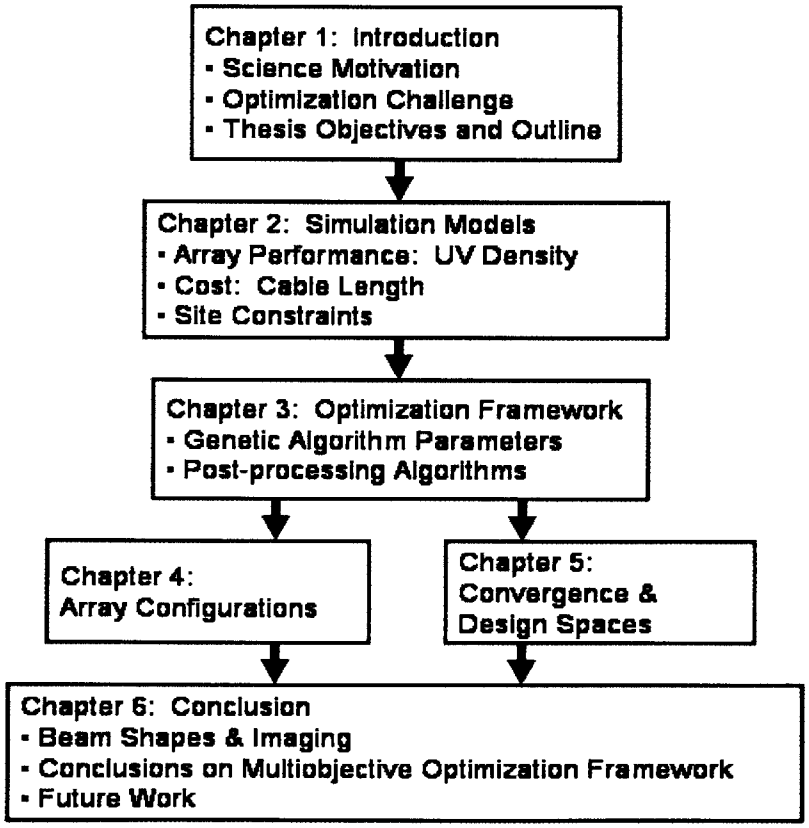


Figure 1-4: Thesis Roadmap

To perform this study, a genetic algorithm was incorporated as mentioned above,

but models were also developed for the LOFAR-specific case and changes were made to the standard Genetic Algorithm Toolbox for Matlab [15].

Chapter 2 will discuss the in depth modelling done for array performance and cable length calculations. The array performance metric has been adopted from Cornwell [5]. Cornwell's algorithm goes as $N^4_{stations}$ in computational time, where $N_{stations}$ is the number of stations. He performed simulations for configurations up to twelve stations. Anything larger became computationally time consuming. This study needs to consider number of stations one or two orders of magnitude larger making the $N^4_{stations}$ algorithm for the performance metric very costly. An approximate method having a computation time of $N^2_{stations}$ was developed and will be discussed further in Section 2.4.1. Cable length determination will be discussed in Section 2.4.2. Site constraints will be discussed in Section 2.4.3.

Chapter 3 will discuss the optimization framework which has been developed. The logic behind the standard genetic algorithm routines [10] and the modifications made to the framework for multiobjective array configuration will be explained. With the models developed in Chapter 2 and the optimization framework laid out in Chapter 3, simulations were run to show the effectiveness of the method.

Chapters 4 & 5 will show results from the LOFAR-specific study. Chapter 4 will give results for unconstrained and constrained configurations. Chapter 5 will give convergence information and comparisons of Pareto fronts across different design and site parameters.

Chapter 6 will summarize the key results from all optimizations. Examples of beam shapes from the optimization runs for the LOFAR-specific case are given. Suggestions for future work will be offered along with difficulties encountered throughout the process of writing this thesis.

1.4 Chapter 1 Summary

Chapter 1 gave the background and motivation behind the need for a framework for multiobjective optimization in the development of radio telescope arrays. Different

types of arrays and possible science concerns are discussed within this thesis using LOFAR [14] specific objectives and constraints as a case study to test the framework. Site constraints, a pivotal part in the design of any ground-based system, is a major part of this study as well, with results and comparisons of configurations to come in subsequent chapters.

Chapter 2

Models for Ground-Based Radio Telescope Arrays

This chapter describes the models developed to assess array performance, cost, and site constraints. Assumptions, design parameters, design objectives, and constraints are discussed.

2.1 Modelling Assumptions

Several assumptions are made in the models for simplification. These are:

- The design domain is circular and flat, neglecting Earth's curvature so that the problem can be done in two dimensions instead of three.
- Stations are treated as points, not extended apertures. This can be done if the station size is much smaller than the overall site size.
- Site constraints only affect station placement, not cable laying. This was done for ease in computation.

The purpose of these models was to demonstrate the use of the optimization framework for competing objectives subject to site constraints. The focus of the models was mainly to reduce computation time. For more rigorous studies, more advanced algorithms can be developed.

2.2 Design Parameters

The key parameter used in the simulations is the number of stations, $N_{stations}$, which is related to the number of samples in the uv plane by the simple relationship in Equation 2.1.

$$N_{uv} = (N_{stations})(N_{stations} - 1) \tag{2.1}$$

This number has a great effect on both the cable length and the number of uv points in the visibility plane. The number of stations itself can be changed to be a design variable to be optimized upon. It may be possible that maximum values of performance or cable length is limited by the number of stations. This is an interesting problem in itself, but has not been chosen as a part of this study specifically.

2.3 Design Variables

The design variables for the simulations are the xy positions of the stations in the antenna plane. Thus, the number of design variables depends on the design parameter of number of stations, $N_{stations}$, and is just twice this number. The xy positions are varied throughout the simulation to obtain an optimal set of solutions.

Initialization of the xy positions of the stations come from already known geometric topologies, shown in Figure 2-1, including Kogan's circular arrays [12], Keto's Reuleaux triangles [11], and the VLA [20]. Non-geometric arrays, created by randomly placing stations in the xy plane, have also been included for comparison and are shown in Figure 2-2. All of these initial seeds adhere to the constraint of placing stations within a diameter of 400 kilometers, to corresponding to LOFAR [14] size parameters. These seeds were chosen to search for improvements on already known well-performing configurations. Results will show where these designs lie in the optimized objective space.

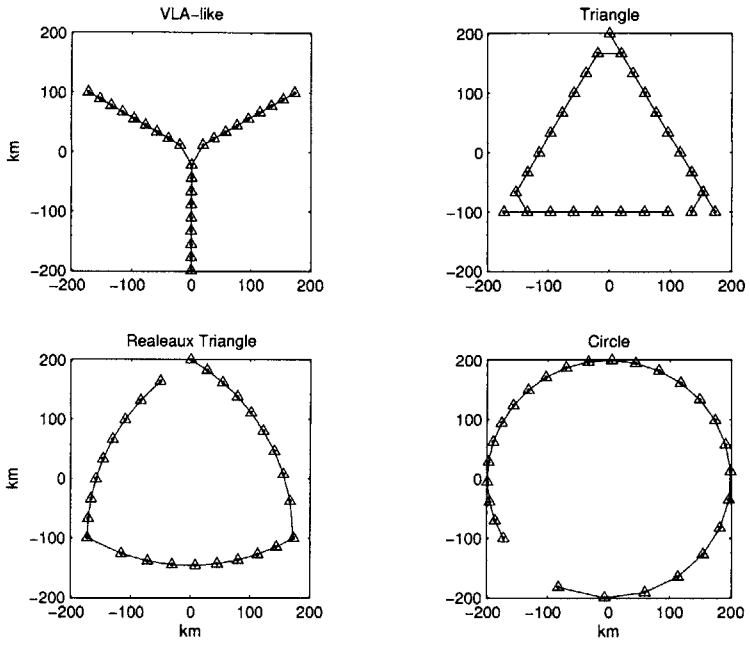


Figure 2-1: Geometric initial population seeds. Shown here with 27-stations.

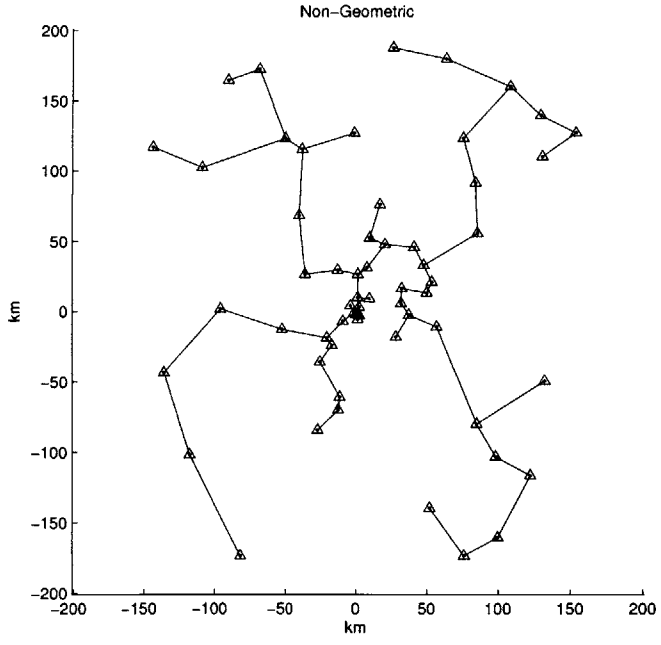


Figure 2-2: Non-geometric initial population seeds. Shown here with 60-stations.

2.4 Design Objectives

2.4.1 Array Performance

Cornwell [5] proposed maximizing the mean distance between uv points as a way to eliminate redundant points in the uv plane. Elegant as his analysis is, it is very computationally expensive and thus not well suited for optimizations on the order of 27-160 stations, which is being considered here. A lower-ordered algorithm has been developed, similar to Cornwell's, to calculate the distribution of uv points for a much larger number of stations.

The method developed for these simulations, referred heretofore as the UV Density Method, first calculates a nominal grid for the number of stations. The number of stations directly determines the number of uv points, given in Equation 2.1. The nominal grid consists of uv points placed according to the desired radial distribution, and then spaced apart equally in azimuth at each radius. At each radius, the azimuthal component of the uv points in the nominal grid is given a slight random offset to induce another level of non-redundancy.

Nominal uv Grid: get_grid

The nominal grid is calculated once per simulation and is used as a benchmark for the statistical evaluation of each design produced throughout the optimization. Figure 2-3 shows a uniform distribution.

Since the number of stations, $N_{stations}$, is fixed for an optimization run, the number of uv points can be calculated for that run from Equation 2.1. Knowing the number of uv points, N_{uv}, the maximum radius of the xy plane, *xygrid* (thus the maximum radius in the uv plane as well by a scaling factor associated with the wavelength of operation, λ), and that a flat distribution is desired, are the only things necessary to compute the grid.

The basic idea of acquiring a flat distribution in the uv plane is to have all uv points roughly the same distance from their neighboring uv points. This is equivalent to having all uv points occupy an equal area in the uv plane.

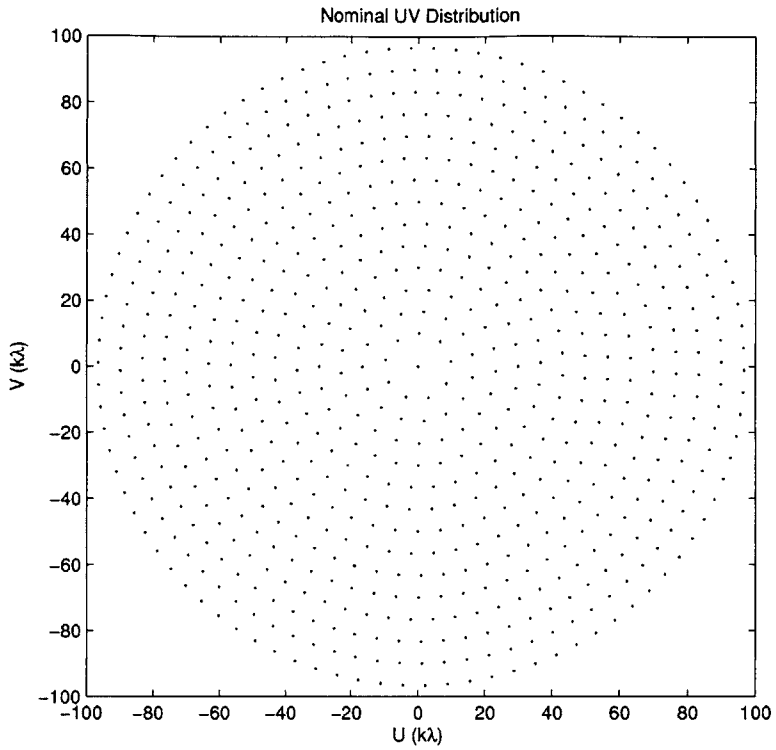


Figure 2-3: Nominal uniform uv distribution for a 27-station (702 uv point) configuration.

To start the total area is calculated from *xygrid* in Equation 2.2.

$$A = \pi \left(\frac{xygrid}{\lambda} \right)^2 \tag{2.2}$$

where A is the area, *xygrid* is the maximum radius, and λ is the observing wavelength. Dividing the total area by the total number of uv points for the simulation yields the area occupied by a single uv point,

$$A_1 = \frac{A}{N_{uv}} \tag{2.3}$$

where A_1 is the area for one uv point, A is the area calculated in Equation 2.2, and N_{uv} is the number of uv points. The grid is developed on a circular grid, so polar coordinates are used throughout the development of the grid. This leads to an assumption that needs to be made about the shape of A_1. In a rectangular coordinate system, A_1 would just be a square with equal sides. Here each area consists of an

inner and outer arc length (s_i and s_o) and straight edged sides (Δr_1 and Δr_2) which are equal. It is assumed that the grid cell is approximately a square with the central arc length, s_c, being the width of the square. Figure 2-4 shows a grid cell.

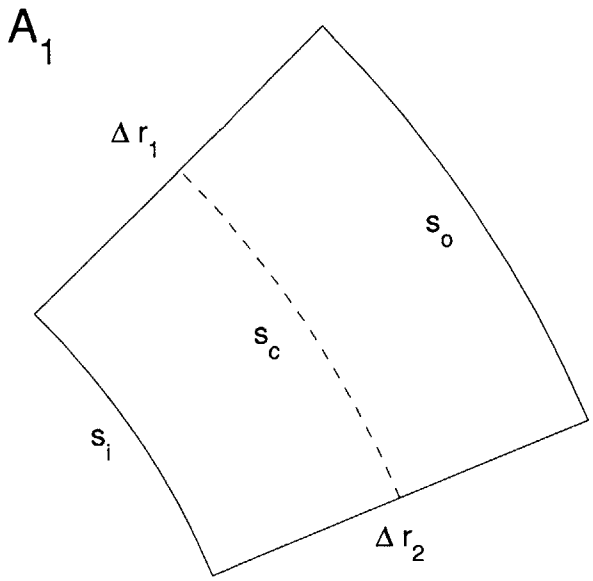


Figure 2-4: A_1 grid cell in the uv plane.

So far just A_1 has been calculated from *xygrid*. What needs to be calculated is the radial and azimuthal components of the nominal uv points. Assuming that A_1 is roughly a square allows Δr to equal s_c and thus be the $\sqrt{A_1}$. Knowing Δr is the essential component in constructing the nominal flat grid because all cells will have the same Δr. The number of annuli in the grid is derived in Equation 2.4.

$$N_{annuli} = \frac{xygrid}{(\lambda)(\Delta r)} = \sqrt{\frac{A}{\pi}}\sqrt{\frac{N_{uv}}{A}} = \sqrt{\frac{N_{uv}}{\pi}} \tag{2.4}$$

where N_{annuli} is the number of annuli, *xygrid* is the maximum radius in xy space, λ is the observing frequency and Δr is the $\sqrt{A_1}$ and the length of a side of a grid cell. From Equations 2.2 and 2.3 a form is derived which depends only on the number of uv points, N_{uv}. From this the radial positions of the centers of the annuli are calculated and stored in a vector $R_{positions}$. This vector is used in a loop to go through each annulus and compute the number and azimuthal positions of uv points in that

annulus. Using the assumption that Δr equals s_c and knowing the radial position of the annuli, the number of uv points per annuli can be calculated. Knowing the $R_{positions}$ of the annuli, the arc length (perimeter) can be calculated from Equation 2.5.

$$S_{annuli} = 2\pi R_{position} \tag{2.5}$$

where S_{annuli} is the perimeter length at the annulus corresponding to the current $R_{position}$. Dividing the current S_{annuli} by Δr (which equals s_c) yields the number of uv points per annulus. These can be easily laid out evenly along each annulus and a random offset can be added between different annuli to add another level of non-redundancy.

The grid has now been constructed except for a problem that occurs at the center. Here the assumption does not hold that grid cells can be approximated as squares because the shape of a cell is now a pie slice, instead of what is seen in Figure 2-4. To include points which lie in this area a point is placed at the origin of the nominal uv grid to account for uv points which may lie near there in an actual grid.

Since the optimization works in rectangular coordinates and the grid is constructed in polar coordinates, a simple transformation converts the grid and stores the rectangular coordinates for use in calculating the metric.

UV Density Method for Minimum Sidelobes: uvdens_circ

The nominal grid is calculated once and it is then called to assess the actual uv distributions from the configurations created throughout the genetic algorithm. Actual uv distributions produced throughout the optimization are calculated from Equations 2.6 & 2.7.

$$u_{i,j} = \frac{x_i - x_j}{\lambda} \tag{2.6}$$

$$v_{i,j} = \frac{y_i - y_j}{\lambda} \tag{2.7}$$

where $i \neq j \;\; \& \;\; i, j \in \{1, 2, \ldots, N_{stations}\}$.

The autocorrelation points (located at the origin in the visibility plane) are removed from the set. An actual uv point from a design is associated with the nearest uv point in the nominal grid using a nearest neighbor algorithm from Matlab called *dsearchn*. If a uv point is associated with a nominal grid point, the nominal grid point is considered filled. $N_{uvactual}$ is the number of nominal grid points which have an actual uv point associated with them.

The metric is defined by counting all nominal grid points which have not been filled ($N_{uvactual}$) and subtracting them from the total number of uv points (N_{uv}). This difference is divided by the total number of uv points (N_{uv}) to give a percentage of nominal grid points that are not filled. The form of this metric is given in Equation 2.8.

$$M = \frac{N_{uv} - N_{uvactual}}{N_{uv}} \tag{2.8}$$

It is desired to have all of the nominal baselines filled, thus having the array performance metric, M, be zero. This is not physically possible, due to the nature of the cross correlation function, but it is the goal. This metric has been chosen for sampling the Fourier plane as completely as possible. The smaller the deviation from the nominal case, the better the design appears in the optimization.

2.4.2 Cable Length Minimization: minCable

For any configuration of stations, there is an analytic solution to the minimum cable length problem (also known as Steiner's problem in graphs [8] or minimum spanning trees). The Single Linkage algorithm [17] is used in the simulations to obtain a solution to this problem. Steiner's problem in graphs [8] also addresses the issue of minimum cost cable laying. For now, without further information on the relationship between cable length and cost, it is assumed that each unit length of cable is of uniform cost, thus simply minimizing the cable length is sufficient. Information providing cable laying costs, or something comparable (e.g. site specific cost maps) can be used to

minimize cost instead of length.

The algorithm for minimizing cable begins by computing all of the Euclidean distances between all pairs of stations and sorting them in order from shortest pair to longest pair. The number of distance pairs is given by the simple relationship in Equation 2.9.

$$N_{pairs} = \frac{(N_{stations})(N_{stations} - 1)}{2} \tag{2.9}$$

where N_{pairs} is the number of pairs made and $N_{stations}$ is the number of stations. Thus measured and sorted, the Single Linkage algorithm begins by creating a connection between the shortest pair. This shortest pair becomes a cluster and is added to the list of N_{pairs}, placed at the rank of $N_{pairs} + 1$. The distance of cable made is recorded and the process continues in this manner of sorting and creating the next shortest connection until all stations are connected.

This method ensures the shortest overall cable length, but solutions need not be unique. There are times when the distances between two groups of points is identical. Either connection can be made and the process will continue, still resulting in a minimum cable configuration. Table 2.1 includes analytic results and algorithm results for the geometric initial seeds given in Figure 2-1. As can be seen from the table, the values compare very well. Differences occur in the Reuleaux triangle and circle because the analytic solutions are based upon arcs while the algorithm creates straight lines. It should also be noted that the cable structure will never have a closed loop, so it should not be surprising that the cable length for a circle is not $2\pi r$.

Shape	Analytic (km)	Computed (km)
VLA	610.3	610.3
Triangle	1000.7	1000.7
Reuleaux Triangle	1025.5	1025.1
Circle	1130.1	1128.8

Table 2.1: The minimum cable routine is benchmarked against the geometric initial seeds shown in Figure 2-1.

2.4.3 Site Constraints: site_con

Site constraints are handled in two places in the genetic algorithm. Since site constraints may go from non-existent, to sparse, to a large portion of the xy plane, it is difficult to tell if the chosen initial seeds will correspond to them or not. So first they are dealt with right after the initial seeds of the population are set. This may be time consuming depending upon how much the site constraints are violated, but it is done once initially for this reason. After the initial seeds are dealt with and reconfigured to adhere to the site constraints, the stations will always fall into allowed zones unless mutation places them in a "badzone". That is why the second place that site constraints is dealt with is in the mutation routine (Section 3.2.6). There the computation is much less, because only the newly mutated points need be checked to see if they violate the site constraints. Since mutation usually occurs in a very small fraction of the overall population, it is not that computationally costly.

The algorithm for dealing with site constraints begins by introducing a new site constraint specific parameter, *gridstep*. These constraints are assumed to be on a square grid. The new parameter specifies the size of a side for one square grid cell in kilometers. The site masks can be set up in any way the user desires; this way was chosen to follow the methods for LOFAR [14] site optimization.

Next the "badzones" are set up. These refer to the areas where stations cannot be placed. For the purposes of this study, badzones are allowed to have cable laid through them. Defining the badzones can be done by either handpicking the zones, or reading them in from a digital map. Once they are defined, all of the stations are checked one by one to see if they fall into a badzone. If they do not, they are left alone. If they do, they are placed into a new spot randomly (similar to how mutations are done). It is possible that the new throw will result in placing the station in a badzone once again, so this is checked and the process repeated until the current station that is being worked on has landed in a safe area. This method is different from that of Boone [3]; he places stations on the borders of the badzones that they fell into. It was determined that there may be times where the complexity of the site

masks may be such that randomized throwing will handle the problem more easily than placing the station on the badzone borders. Once all stations are in safe zones, the optimization can continue.

2.5 Chapter 2 Summary

Chapter 2 discusses the parameters, variables, objectives, and models used in the multiobjective optimization simulations which have been done for this thesis. The design parameter for this study is the number of stations. The design variables are the xy positions of the stations. The metrics of array performance and minimizing cable length were motivated by a study for the design of LOFAR [14]. Site constraints limit the placement of stations on the ground, and how these constraints are handled is explained.

Chapter 3

Optimization Framework

3.1 Genetic Algorithm

A genetic algorithm [24] with tournament selection [15] has been implemented for the framework of the optimizations. As described above, a genetic algorithm is a good technique for nonlinear and multiobjective optimizations because it is a heuristic technique and is population-based, working on more than one solution at a time. Selection is a key aspect to the operation of genetic algorithms. Tournament selection is chosen due to the fact that no weighting function between the objectives is required. Weighting can occur in post processing, independent of the optimization itself, and will be discussed further in this section.

Multiobjective optimization introduces a trade off between the design objectives, producing an optimal (non-dominated) family of solutions which is a subset of the objective space. Figure 3-1 shows a graphical representation of a two-dimensional (two objective) objective space with important features outlined further below.

The subset of optimal solutions is all of the non-dominated solutions which lie on or near the Pareto front [24]. The non-dominated solutions are defined by the set of feasible solutions such that there are no other solutions in the objective space for which one design objective can be improved without reducing the optimality in another design objective. The Pareto front is a theoretical front made up of full convergence in the non-dominated solutions. In this framework, all solutions which

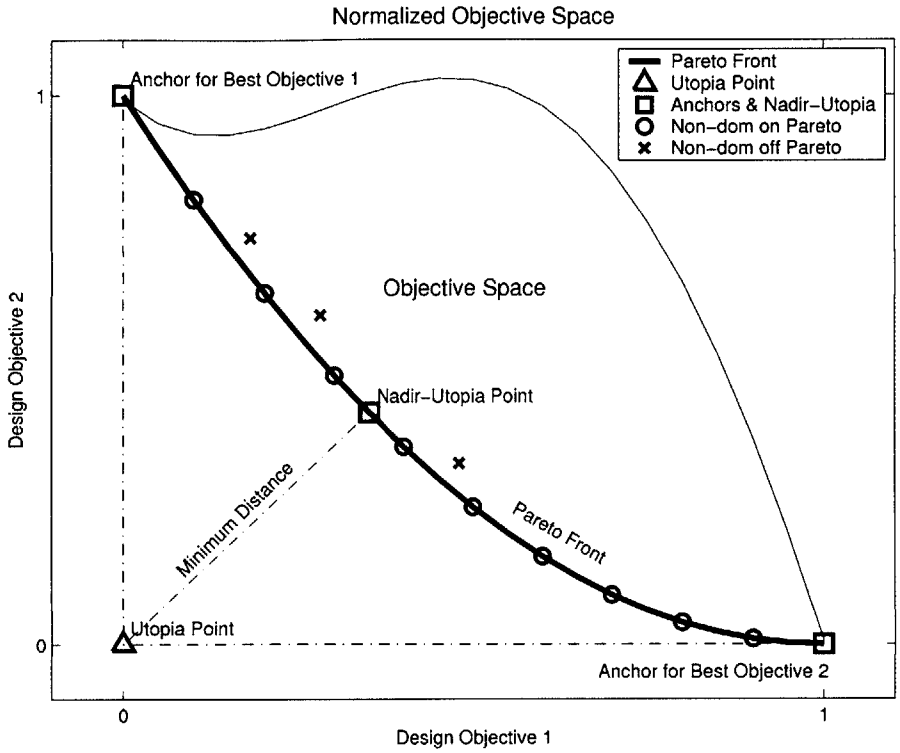


Figure 3-1: Normalized objective space. Anchor and nadir-utopia points are shown as squares. The utopia point is shown as a triangle. Non-dominated solutions are shown as both those that lie on the Pareto front (circles) and those which lie off the front (crosses). This objective space assumes minimization of both objectives.

are non-dominated, even those which may not lie on the Pareto front, are presented. Due to the nature of genetic algorithms, full convergence in the Pareto front has not been proven.

There are two types of solutions on the Pareto front which are of special interest: the anchor points and the nadir-utopia point. Anchor points are defined as the points on the Pareto front which contain the best achievable value of one of the design objectives. In this case, since there are two objectives, there are two anchor points. The nadir-utopia point is found by first normalizing the axes with respect to the anchor points. The anchor points are located at opposite corners of the normalized objective space, in this case they have both been set to a value of 1, since both objectives are to be minimized. The utopia point, defined as the theoretical optimum (not achievable in this case), lies at the corner of the objective space representing improvement in

both design objectives (at the origin for this problem). The nadir-utopia point is defined by the point on the Pareto front which is the minimum Euclidean distance, D_{NU}, (refer to Equation 3.1) from the utopia point in the normalized objective space.

$$D_{NU} = \min \sqrt{\left(\frac{J_1 - J_1^*}{J_{1norm}}\right)^2 + \left(\frac{J_2 - J_2^*}{J_{2norm}}\right)^2} \tag{3.1}$$

where $\{J_1^*, J_2^*\}$ represents the utopia point, $\{J_1, J_2\}$ represent points in objective space, and $\{J_{1norm}, J_{2norm}\}$ are the normalizations of the axes on the objective space plot. Different relative weighting of the objectives can be implemented by different scalings of the axes, resulting in different nadir-utopia points.

3.2 Genetic Algorithm Routines

The standard genetic algorithm [10] has been modified for this problem. New design variables may be made which are not the xy positions of the stations. Since the number of stations, $N_{stations}$ directly relates to the number of design variables currently, it would be easier to wrap a second optimizer around the current one. For instance if the value of $N_{stations}$ itself were a design variable, the number of design variables would not be constant throughout a run, making it very difficult to keep track of solutions throughout the simulation. Wrapping another optimizer around the current one would alleviate this problem, but would increase the time of optimization considerably.

The algorithms used in this case study are explained below. Modifications have been made to nearly all standard genetic algorithm operators [10] such as selection, crossover, and mutation. Brief explanations of these modifications will also be given below.

3.2.1 Main: GAuvdens

GAuvdens is the driver program for the optimization code. The acronym stands for "Genetic Algorithm UV Density Method". Parameter set up and all higher level

functions reside in the driver. If site constraints are to be considered, the slightly modified driver of GAsite can be used. Both are included in Appendix B.

Parameters

The design parameter, the number of stations ($N_{stations}$), is defined at the start of GAuvdens. Genetic algorithm parameters [10] are defined next. These parameters include the number of generations (*ngen*), population size (*popsize*), crossover rate (*xrate*), mutation rate (*mrate*), and elitism rate (*elite*).

Optimization constants follow. The diameter of the physical space (*xygrid*) is given in kilometers. This number is only used for scaling in the station (or xy) plane and has no bearing on the optimization. It is set to 400 km to coincide with LOFAR [14] site constraints. The wavelength (λ) of the signal is just a scaling factor, scaling in the baseline (or uv) plane.

The site constraint parameters *percent* and *gridstep* are included for optimization runs with site constraints. *Percent* refers to the percentage of badzones that cover the xy plane and *gridstep* determines the size of the badzones in kilometers.

Initialization

Genetic algorithms require initial guesses to determine the first generation of the population. These guesses are important because they determine the future of the evolution of the overall population. If the population is too narrow initially, interesting solutions may never be found that otherwise could have been with a more diverse population. Many times there is no good guess of what a final solution may be, so different initial conditions need to be tried to explore the objective space as thoroughly as possible. The initialization of the populations takes place in the subroutine **init**, discussed earlier in Section 2.3. Both geometric and non-geometric arrays are used.

The nominal grid from Section 2.4.1 is also initialized here and stored for use in the selection routine. This routine is very time consuming, and performing it once at the beginning lowers the overall computational load required.

If site constraints are being considered, the initial population is put through

site_con and adjusted for the corresponding site constraints. This will often break the symmetry of the highly geometric shapes, making them slightly randomized, but site constraints make it difficult for highly geometric designs to exist in reality.

Optimization

The evaluation of the performance metrics, discussed in Section 2.4, takes place in the selection routine **select_uvdens**. Here the tournament selection [15] takes place. The array configurations are evaluated and the fittest survive and continue in the evolutionary process. Further details of how the tournament selection works will be discussed in Section 3.2.2.

Throughout the optimization the best designs need to be saved for use in enhancing the population, as well as defining the Pareto front [24] of solutions. The Pareto optimal set of solutions is used in the elitism subroutine, discussed further in Section 3.2.4, to reintroduce anchor solutions back into the population, thus expanding the Pareto front.

Up to this point, everything has been done once to set the optimization in motion. This first generation is considered the "parent" generation from which the remaining optimization now starts. A 'while' loop is created that cycles through the genetic algorithm operations until the maximum number of generations (*ngen*) is reached (alternatively, lack of change in the Pareto front could be used). Now the operations of elitism (Section 3.2.4), crossover (Section 3.2.5), and mutation (Section 3.2.6) occur, which evolve the population and improve the design in all of the objectives considered.

After the evolutionary process has occurred once, through selection, Pareto selection, elitism, crossover, and mutation, it is time to move on to the next generation. Just like in biological evolution [6], the genetic algorithm has kept the fittest of the population and passes on those traits to the next generation, to keep improving the design. The process is then repeated until the final selection takes place. The best non-dominated solutions are always kept in memory.

Post-Processing

Due to the large number of array configuration which are evaluated throughout one optimization run, not all of the station placements of the configurations are saved. What information can be saved to offer the most information, but also is a manageable amount of data? As the performance metrics are evaluated they can be stored without much memory. Each configuration has anywhere from 27 to 160 stations (in accordance with LOFAR parameters), each with an xy position, which is a lot of data to store if 100,000 designs are being evaluated throughout one optimization run. Each configuration has two objectives (in this study) for which metrics can be stored, thus reducing the amount of data by 27 to 160 times.

Thus the values of the design objectives can be stored to create information on the optimization run to show population evolution and convergence information (Section 3.3.2) which plots the figures of merit, or design objectives as stated above. The designs which are most interesting are the Pareto optimal set of solutions. In the Pareto filter (Section 3.2.3) subroutine, these solutions station placements are stored to be looked at in the future by the plotting routines (Sections 3.3.3 & 3.3.4).

The array configurations that are most interesting are the non-dominated solutions. The xy positions of these configurations are stored and continuously updated as new configurations dominate old ones. This makes for a much more manageable amount of data.

3.2.2 Selection: select_uvdens

There are many forms of selection criteria that can be used in genetic algorithms. In a framework for multiobjective optimizations, weighting functions will usually be required. Weighting functions are problematic when one of the multiple objective metric values is improved much more quickly than the other, skewing the distribution of the weighting function. Tournament selection requires no weighting of the objectives a priori. It is a much simpler method, both in principle and in computation. The tournament selection function written here takes in the xy positions of the

stations, and the grid information from **get_grid** (Section 2.4.1) used in evaluating the performance metric.

First, an index array p is assigned which stores the integers from 1 to the population size, *popsize*, used in the current optimization run. This index array will be used in the future to randomize the order of the members of the population during the selection process. This is important to do, and is done not only in this routine, but in others, so that traits are allowed to propagate throughout the population, and are not isolated in one place. A new array, *xynew* is also initialized which corresponds to the same size as the array which stores all of the current configurations and will be used as a dummy array to store the winners in the selection process which will then be passed back to **GAuvdens**.

Once the initialization processes are finished, a loop is created which steps from 1 to *popsize* by twos. It steps through by twos because in this case there are only two objectives. This is a very important point which comes up in the design of the code and pertains to the number which is assigned to *popsize*. *Popsize* must be divisible by the number of design objectives. Here there are only two, so *popsize* needs to be an even number. Now the index array p is used to pick out at random two array configurations at a time and evaluate both of their design objectives: array performance and cable length. These values are temporarily stored in two types of dummy variables, each with two members. One group is made up of (P_1, P_2) for the two performance metrics which were just evaluated, the other group is made up of $(cable1, cable2)$ for the two cable lengths which were just calculated. Now the tournament can occur between the two current designs. First the two array configurations are compared in the performance metric. Since a lower numeric value is desired for the performance metric, if $P_1 < P_2$, the first configuration wins; if $P_1 > P_2$ the second configuration wins. The same is done with $cable_1$ and $cable_2$, with a lower numeric value being better and thus the winner. The winner of each competition has its xy station placements stored in *xynew* and its corresponding performance metric and cable length stored in two arrays named Pi and *cablei*. The winner of the performance metric competition is stored in the odd index, the winner

of the cable length stored in the even index. In this way the arrays can be filled by stepping by twos.

In the competition, one of the array configurations wins the performance metric competition and is thus saved once, and the other array configuration wins the cable length configuration and is also saved, or one of the array configurations wins at both competitions. Most of the time the former case occurs, but it is when the latter case occurs that the selection routine does its work, because it copies over the very good array configurations twice (once for performance and once for cable length) thus increasing their chances to both survive and pass on traits to the next generation as the process of evolution continues on. These configurations are the ones which will advance the nadir-utopia part of the Pareto front, while the very best in the single objectives will expand the Pareto front near the anchor points.

3.2.3 Pareto Filter: pareto

Pareto optimality is defined as an improvement in one objective that can only be achieved at the expense of at least one other objective [24]. Non-dominated solutions are the ones found throughout an optimization. The Pareto front itself is a theoretical curve which corresponds to full convergence in the non-dominated solutions.

The algorithm for determining if new solutions are Pareto optimal also needs to remove old non-dominated solutions which become dominated by the new solutions. This can be done by creating a superset of solutions that includes all members of a new generation with all members of the previous non-dominated set. Then the Pareto filter can evaluate every solution against all others and the ones which are non-dominated become the new set of non-dominated solutions. This process does not require much computation and can be easily done from generation to generation, resulting in the best non-dominated set of solutions at the calculation of the final generation.

3.2.4 Elitism: elite

Elitism is used to expand the Pareto front near the anchor solutions. It assists in keeping the population more diverse throughout the Pareto front. The Pareto front has a tendency to clump in certain regions as families of solutions begin to emerge. Elitism, along with other possible operators (such as selective mating [10]), can assist in reducing this tendency to clump.

One of the optimization parameters is the rate of elitism. This determines what percentage of the population should be made up of the anchor solutions. This number is low (in the range of a few percent) because if it were too high then the optimization would be biased too much towards the anchor solutions. The elitism algorithm searches the current non-dominated set of solutions and picks out the correct number of anchor solutions based upon this number. The anchor solutions are then copied and placed back into the population randomly, replacing other solutions so as to keep the *popsize* constant. The storage vectors for xy positions and design objectives is updated accordingly.

3.2.5 Crossover: xover

The crossover routine is one of the essential genetic algorithm operators. It is analogous to mating and exchange of genetic material [6]. Crossover takes two configurations (or sometimes several, this is where the analogy to mating breaks down, but is still computationally feasible) and exchanges pieces of the information strings between the two. For the framework of array configuration optimization, this is the swapping of xy stations between different configurations. The optimization parameter of crossover rate controls what percentage of members of the population are allowed to exchange station placement information. This parameter is usually fairly high (between 80 to 95 percent). It is also important to decide how much information is to be exchanged. There are many methods for crossover [10]. Here, two configurations are taken and "cut" in random places and the information is exchanged one-to-one.

An interesting issue arises with relation to crossover and the type of problem

that needs to be solved here. Discussed further in Section 6.3 will be the concept of destructive mating, in which mating may actually produce poorer designs more often than it produces better designs. This is due to the high nonlinearity in the design objectives with relation to the design variables (xy placement of the stations). A small change in a small percentage of the stations of a configuration may cause drastic change in the design objective metrics.

3.2.6 Mutation: mutate_circ

Mutation is the operation by which searching the objective space is expanded upon. Mutations allow for new solutions to appear, in this case by introducing new station placements into the configurations. The mutation rate is very critical in determining the outcome of the solutions. If it is too high, the genetic algorithm may operate in a Monte Carlo fashion, constantly sampling the objective space but not performing any real optimization because the selection operator is undermined. Convergence may never be achieved in this case. On the other hand, if the mutation rate is too low the solution may converge prematurely. The trade off between high and low mutation rate is a difficult one to deal with in genetic algorithms. Often (as was done here) several runs with varying mutation rates are required to determine the best one for the specific problem.

As was discussed earlier in Section 2.4.3, site constraints are also handled in the mutation routine. Every time that a station is mutated to a new xy position, it is checked to make sure that it has not landed in a badzone. If it has, it is mutated again until it has landed in an allowable zone.

3.3 Post-Processing Routines

Post-processing routines are very critical in evaluating the results of the optimization. Here is where the design of the multiobjective optimization framework can be graphically represented to compare objectives and relationships between different configurations.

The post-processing routines can also be modified to include other relationships that future users may desire to develop on their own and include into the framework. Current post-processing algorithms show plots of convergence in the design objectives, objective spaces with evolutionary information and Pareto fronts, and layout designs for non-dominated station configurations. As objectives may change, or new ones get added, the post-processing framework is very adaptive to representing the new information in new ways.

3.3.1 Nominal Grid: GRIDplot

A simple routine has been developed which will just take in the nominal uv grid points from **get_grid** and plot them so a visual comparison can be made between the goal that is being strived for and what the actual uv plots look like for the optimized configurations.

The Matlab plotting parameter of *Markersize* can be adjusted to vary the size of the uv points that are being plotted for better visualization. Typically for the 27 station case, which contains 702 uv points, a *Markersize* of size 5 is sufficient. All other number of station cases had a *Markersize* of 1. The axes are in kilo-wavelengths, similar to the uv plots of the array configurations.

3.3.2 Figures of Merit: FOMplot

The figures of merit plots show the comparisons between the design objectives. Plotted are the convergence information for each design objective as a function of generation number and objective spaces.

Convergence plots show the average value of the design metrics across all members of the population as a function of generation number. The objective spaces are color coded to show the evolution of the population throughout the optimization. There is a point for each configuration pertaining to the value of its design metrics (in this study there are two axes). Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%,

blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

3.3.3 Array Configuration w/ corresponding uv Points: uvplot

The non-dominated solutions are plotted at the end of an optimization run, sorted from the best performance anchor solution to the minimum cable anchor solution. The stations are placed down and then the minimum cable plotting function is called. Discussion of the minimum cable plotting function is given in the next section (Section 3.3.4). The badzones from site constraints are given as green squares in the xy plot.

Plots of the corresponding uv distribution are given as subplots. Both metrics of cable length and performance metric value are placed on the plots automatically. It may be necessary to adjust the *Markersize* parameter in Matlab so that the uv points show well, according to the number that appear.

3.3.4 Minimum Cable: minCableplot

The minimum cable plotting routine uses a slightly different algorithm than that of the minimum cable metric calculation, described in Section 2.4.2. Since the cables need to be drawn on the plot the algorithm is slightly more computationally intensive than the Single Linkage Algorithm [17]. Each link that is found must have a connection made between the closest points in the group. This requires that the individual station placements be kept track of with respect to their group, as opposed to clustering them together. The algorithm still gives the same cable length as expected from Single Linkage.

3.3.5 Point Spread Function: PSFplot

This routine reads in the xy positions of a configuration and produces a point spread function. This is done by taking the cross correlation of the xy positions to get the uv positions and then taking the two-dimensional Fourier transform to produce a beam.

This is a computationally intensive process and is not used on every configuration. Different metrics can be derived from point spread function statistics and can be used to compare configurations across the parameter $N_{stations}$.

3.4 Genetic Algorithm Flow Chart

A graphical representation of the general flow of the genetic algorithm framework is given in Figure 3-2.

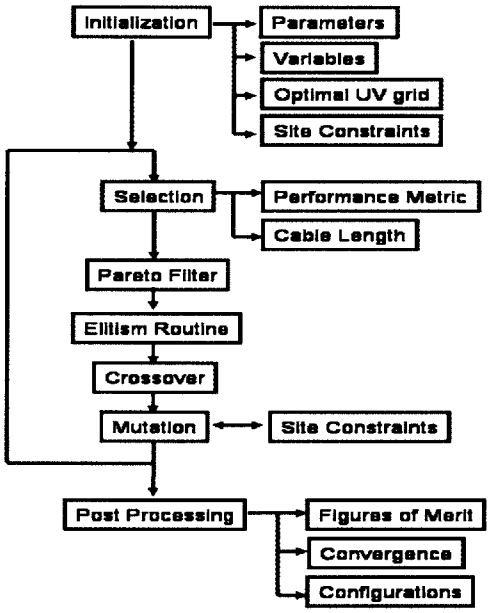


Figure 3-2: Genetic Algorithm Flow Chart

3.5 Chapter 3 Summary

Chapter 3 discusses the use of genetic algorithms in the framework of multiobjective optimization. The genetic algorithm operators are discussed thoroughly in this chapter. Modifications have been made to the standard operators for the specific use of optimizing large radio telescope arrays. Also discussed is post-processing routines developed for the analysis of the results which are presented in subsequent chapters.

Chapter 4

Array Configuration Results

As is discussed in Chapter 1, a uniform distribution in the uv plane corresponds to having minimum sidelobes in the point spread function of the interferometer [5]. For this study uniform uv distribution has been considered as one of the objectives while minimum cable has been chosen as the other. The configurations that are produced from this formulation are presented in this chapter. Chapter 5 will discuss the trends in the objective spaces and convergence information.

In any real world system there are constraints to consider in the design of systems, radio telescope arrays are no exception. Site constraints ranging from mountains, cities, rivers, high interference environments, etc... will affect the placement of stations in the array. Constrained optimizations, as discussed in Chapters 2 & 3, will presented as part of the solution.

An overview of simulation parameters used is given in the next section (Section 4.1). Emerging configurations, consisting of the anchor and nadir-utopia solutions are given in Section 4.2.

4.1 Simulation Parameters

Optimizations were performed for arrays of 27-, 60-, 100-, and 160- stations. Table 4.1 shows the parameters used in these optimizations: the number of uv points (N_{uv}), and the genetic algorithm parameters of population size, number of genera-

tions, mutation rate, elitism rate, and crossover rate. The population is the number of designs which are being operated upon by the optimizer at one iteration. As the number of stations increases, the number of design variables increases accordingly and this usually requires a larger population to maintain diversity throughout the simulation. As the population increases, so does the computation time. The number of generations determines how many iterations the optimizer runs through and is set at a high number to assure convergence in the solution and control the time that one optimization run takes. The mutation rate determines the number of random station position shifts per generation. This allows different parts of the objective space to be explored. The elitism rate determines the number of anchor points which are reinserted back into the next generation to aid in expanding the objective space near the anchor points. The crossover rate determines how many designs from the population are mated per generation. Crossover is the essential operation in genetic algorithms, as it is used to pass on desired characteristics throughout an optimization run (see [24] for an introduction to Genetic Algorithms).

All unconstrained runs were done on a Pentium 4 2000 MHz machine with 1 Gigabyte of RAM. Since the algorithms used in the models go as N^2, the times for simulations also increases by roughly N^2, making it much more computationally intensive as the number of stations increases. The execution time for each simulation is given in Table 4.1.

As discussed in Section 2.4.3, the parameter badzones has been introduced to mask out zones which do not allow stations to be placed, but can still have cable running through them. This part of the study focuses on how the percentage of badzones (*percent*), as well as the size of the badzones (*gridstep*) affect the objective space. The *percent* values chosen were 10, 30, and 50 percent. The unconstrained cases can be seen as the $percent = 0$ case. The *gridstep* sizes chosen were 4, 20, and 40 kilometers. These sizes were not necessarily picked to mirror any known physical objects, but were chosen as a good range in sizes. Badzones are also distributed at random on the xy plane for each simulation run.

For the site constrained part of the study, the population size (*popsize*) and num-

$N_{stations}$	N_{uv}	Population	Generations	mrate	erate	xrate	time
27	$2 * 351$	500	5000	1%	1%	90%	10.1 hrs
60	$2 * 1770$	200	5000	1%	1%	90%	18.3 hrs
100	$2 * 4950$	300	6000	1%	1%	90%	117.2 hrs
160	$2 * 12720$	200	2000	1%	1%	90%	72.3 hrs

Table 4.1: Simulation parameters for genetic algorithm optimization based upon varying number of stations ($N_{stations}$). The number of stations ($N_{stations}$), corresponding number of uv points (N_{uv}), and genetic algorithm parameters of population (*popsize*) and generations (*ngen*) is given. N_{uv} is given as two times the number of independent uv points because of the Hermitian property of the visibility function. Genetic algorithm parameters of mutation rate (mrate), elitism rate (erate), and crossover rate (xrate) are given in percentages. Simulation run times are given in hours and were performed on a Pentium 4 2000 MHz machine with 1 Gigabyte of RAM.

ber of generations (*ngen*) was held fixed throughout all optimization runs, regardless of number of stations ($N_{stations}$). This was done to limit the number of variable parameters in the study. 1000 generations were run with a population size of 100 for all simulations with site constraints. This did not always lead to full convergence (small scale) in the solutions, but this also allows for a study of how convergence is affected by the number of stations. The execution time for each simulation is given in Table 4.2. Runs with a * were performed on a Pentium 4 1700-MHz machine with 256 Megabytes of RAM. Runs without a * were done on the same machine as unconstrained simulations, a Pentium 4 2000 MHz machine with 1 Gigabyte of RAM. For runs performed on the 1700-MHz machine, the time conversion has been done and included in parentheses.

Simulation run times still follow the general N^2 rule that was seen in the unconstrained cases. The main computational load of the site constraints comes in the front end of the computation when the initial seeds are first subject to the site constraints. Since the site constraints are randomly placed down, it is possible that many stations in an initial design fall into badzones. These stations are handled in a way similar to ones later on. The stations which fall into badzones are thrown at random until they do not lie in a badzone. This destroys the initial highly geometric shapes and makes them more randomized to begin with. Discussion of the objective spaces comes in

$N_{stations}$	Population	Generations	*gridstep*	*percent*	time
27	1000	100	4	10	0.47 hrs
27	1000	100	4	30	0.49 hrs
27	1000	100	4	50	0.57 hrs
27	1000	100	20	10	0.42 hrs
27	1000	100	20	30	0.43 hrs
27	1000	100	20	50	0.44 hrs
27	1000	100	40	10	0.42 hrs
27	1000	100	40	30	0.42 hrs
27	1000	100	40	50	0.43 hrs
60	1000	100	4	10	2.1 hrs
60	1000	100	4	30	2.3 hrs
60	1000	100	4	50	2.4 hrs
60	1000	100	20	10	1.9 hrs
60	1000	100	20	30	2.1 hrs
60	1000	100	20	50	2.4 hrs
60	1000	100	40	10	2.0 hrs
60	1000	100	40	30	2.1 hrs
60	1000	100	40	50	2.3 hrs
100	1000	100	4	10	6.5 hrs
100	1000	100	4	30	9.1* (7.7) hrs
100	1000	100	4	50	8.0 hrs
100	1000	100	20	10	6.0 hrs
100	1000	100	20	30	8.6* (7.3) hrs
100	1000	100	20	50	9.7* (8.3) hrs
100	1000	100	40	10	7.7* (6.5) hrs
100	1000	100	40	30	8.1* (6.9) hrs
100	1000	100	40	50	7.5 hrs
160	1000	100	4	10	19.3 hrs
160	1000	100	4	30	24.3 hrs
160	1000	100	4	50	23.9 hrs
160	1000	100	20	10	24.3* (20.7) hrs
160	1000	100	20	30	20.5 hrs
160	1000	100	20	50	23.8 hrs
160	1000	100	40	10	23.0* (19.6) hrs
160	1000	100	40	30	19.7 hrs
160	1000	100	40	50	23.2 hrs

Table 4.2: Simulation parameters for site constraints based upon varying the number of stations ($N_{stations}$), the percentage of badzones (*percent*), and the size of a badzone (*gridstep*) in kilometers. Genetic algorithm parameters of *ngen* and *popsize* remained fixed. All other genetic algorithm parameters remain fixed as well and correspond to values given in Table 4.1.

4.2 Configurations

Sections 4.2.1 - 4.2.4 show the 27-, 60-, 100-, 160-station configurations. Each figure consists of six panels. The top three panels in each figure show the xy station placements. The stations are circled and cable connections are shown with lines. The bottom three panels show the uv coverage corresponding to the xy station placements directly above them. From left to right the configurations presented are those of minimum cable length, nadir-utopia, and maximum array performance. The cable length metric is given in kilometers assuming an overall array diameter of 400 km (the LOFAR specification), and the array design metric, M, is given as the percentage of empty nominal uv baselines with respect to the total number of baselines.

The two objectives of minimizing the cable length and achieving the desired uv distribution oppose each other. The first tends to clump stations together to reduce cable length, while the latter tends to spread stations apart to obtain new uv points. Multiobjective optimization is a way of searching the configuration space for good trade-off solutions. The solutions display a wide Pareto front showing the trade-off between decreasing cable length, and thus decreasing cost, and improving array performance, which will be shown in Chapter 5. Along the Pareto front designs range from VLA-like structures to ring-like structures.

One expects minimal cable anchor solutions to be highly condensed designs, but it should be noted that there needs to be a minimum performance that designers wish to consider, or the minimum cable anchor solution will always be a highly condensed array with all stations clumping to a point. Solutions that have VLA-like configurations are chosen to be the minimum cable configurations for unconstrained configurations. The VLA has been chosen as the minimum cable configuration because designs with lower cable length also tend to have VLA-like characteristics, but do not stretch out to the outer boundaries of the xy plane. In configurations that have site constraints, this limit has been loosened for a number of reasons. As can

be seen from the objective space plots in Chapter 5, many Pareto solutions do not achieve the performance and cable length of the VLA. Another reason is that the constrained configurations were performed with a set population size and number of generations, which may not have allowed for full convergence to be achieved.

In general, minimum cable solutions appeared as slightly randomized VLA-like configurations, while best array performance was achieved by ring-like configurations with inward reaching arms. Solutions near the nadir-utopia point consisted of hybrid solutions of different initial seeds and Reuleaux triangles.

As the number of stations increased, Pareto optimal solutions were more and more like the initial seeds of the population in the unconstrained case. Why is this so? The number of uv points in the uv plane is increasing for a fixed uv plane size. In essence, the density of the uv plane is increasing. As the uv density increases, the gaps in the uv plane become smaller. Smaller gap sizes, along with the large number of uv points makes it difficult for small perturbations from the highly geometric designs to make a large difference in the array performance metric. This creates a diminishing return on increasing the number of stations to fill the same uv space. The highly geometric designs, which have very good cable length qualities, produce very similar results to designs that have small random perturbations and longer cable lengths. In the 27-station case, there were large gaps in the uv plane, thus small perturbations and hybridization aided in improved designs with respect to both metrics. This result suggests that a cost trade-off may exist between adding more stations to a highly geometric design, and moving around a fixed number of stations to create a better uv coverage, a trade-off not considered here. There are other possible considerations as well, such as surface brightness sensitivity and the computational power required to combine more signals from more baselines. New objectives need to be introduced to take these important considerations into account.

For constrained cases, the highly geometric initial seeds were no longer highly geometric once they were filtered through the initial site constraint routine. This makes it difficult to achieve highly geometric low cable configurations along the Pareto front of the constrained cases. Also, as the parameter $N_{stations}$ increases it can be seen

that the site constraints affect the configurations to a greater extent. This can also be seen in results shown in Chapter 5. This may be do to the fact that more stations are randomly replaced during the initial site constraint routine for larger $N_{stations}$. There may be a better way to replace stations that fall into badzones initially which will keep the initial seeds more geometric for the constrained cases. This would involve a more complex algorithm than the one used here, but is a good subject for future work.

In the next several sections are plots of the configurations. They are split into sections by the parameter $N_{stations}$. Further discussion of the objective spaces will occur in Chapter 5.

4.2.1 27-station Configurations

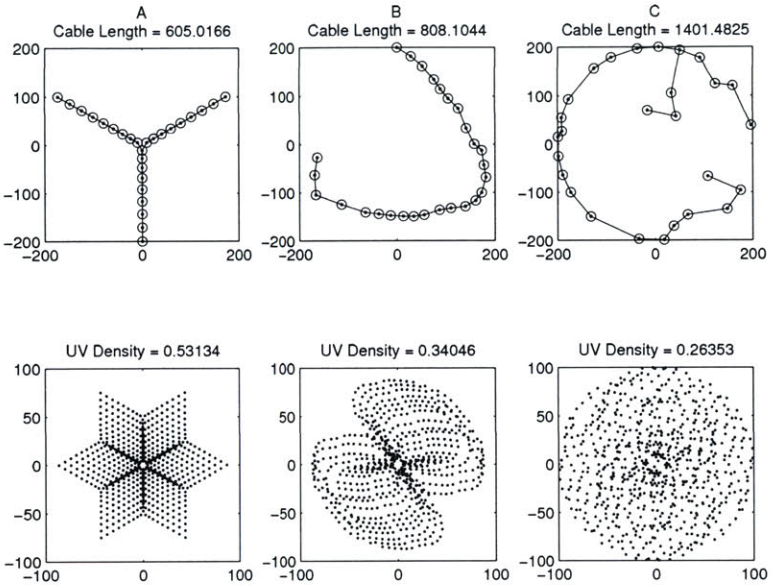


Figure 4-1: Unconstrained 27-station configurations (top) with corresponding uv coverage (bottom). (**A**) Minimum cable (left), (**B**) nadir-utopia (center), and (**C**) maximum performance configurations (right) are shown on Figure 5-8.

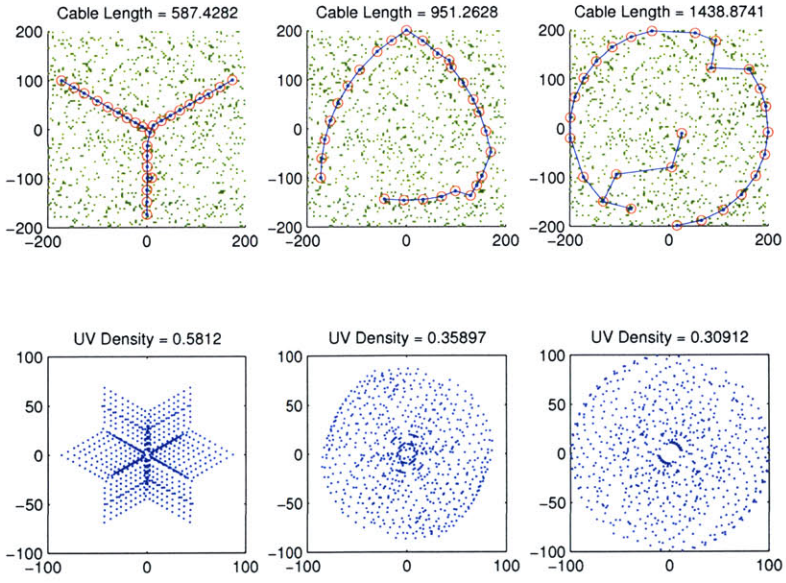


Figure 4-2: 27-station, 4km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

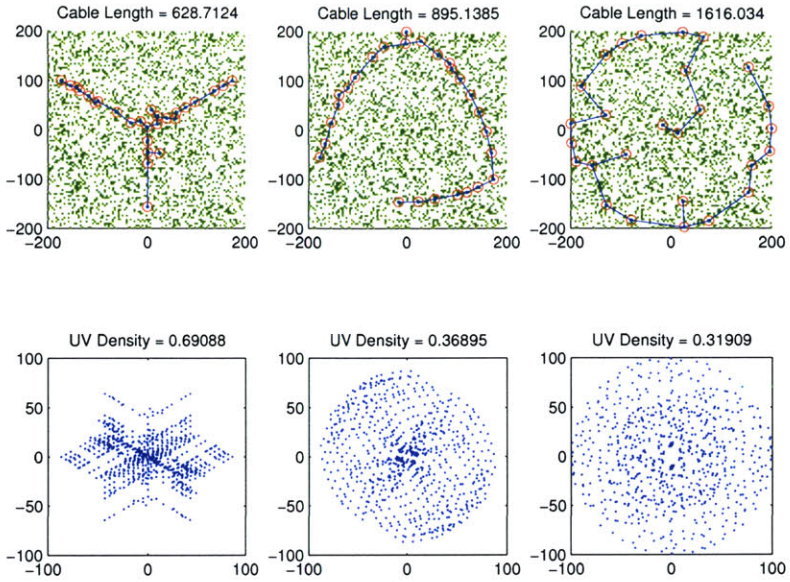


Figure 4-3: 27-station, 4km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

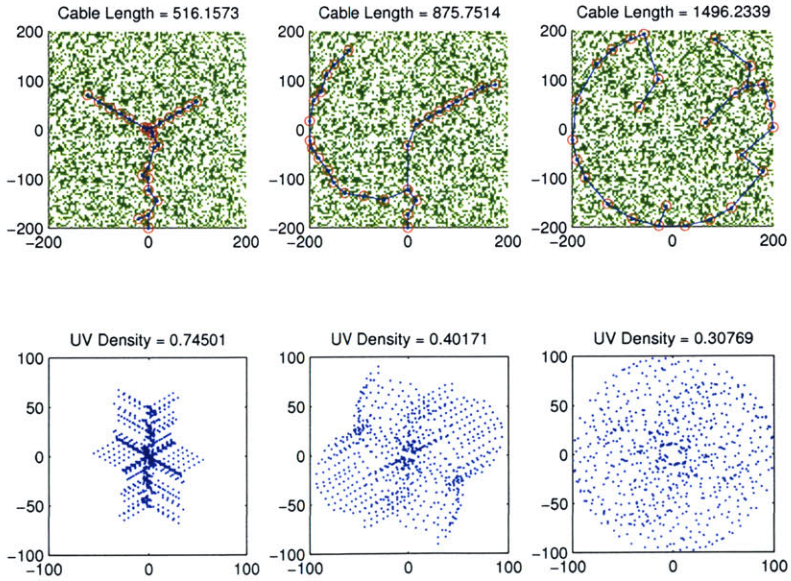


Figure 4-4: 27-station, 4km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

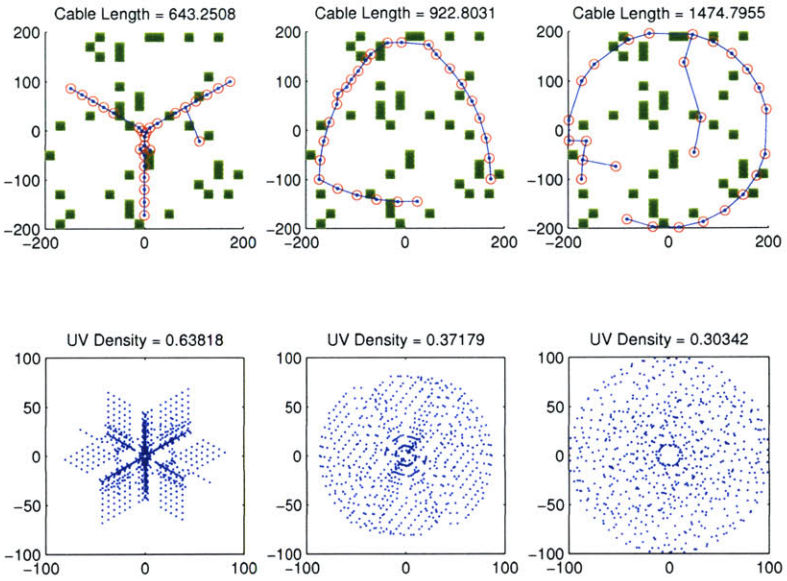


Figure 4-5: 27-station, 20km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

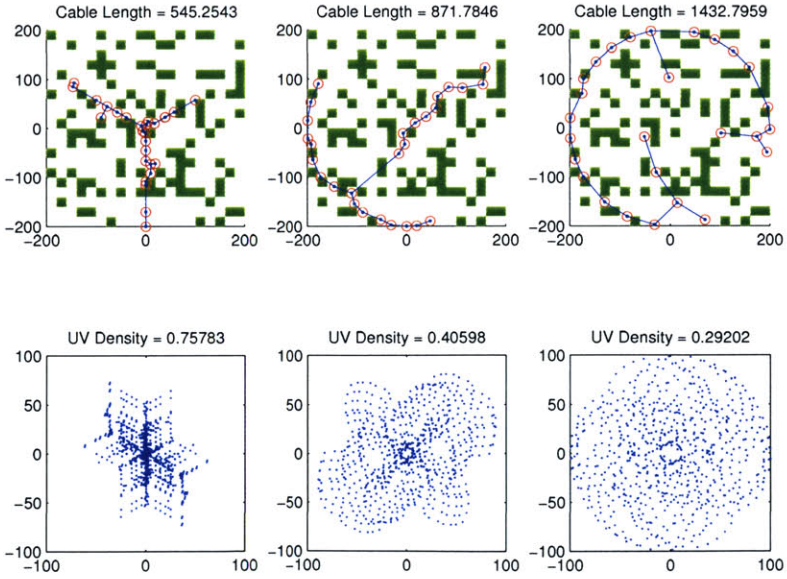


Figure 4-6: 27-station, 20km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

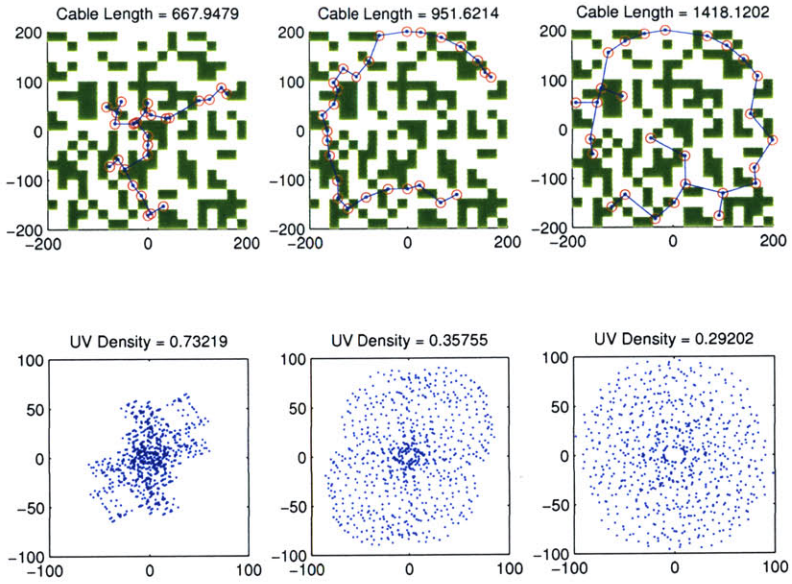


Figure 4-7: 27-station, 20km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

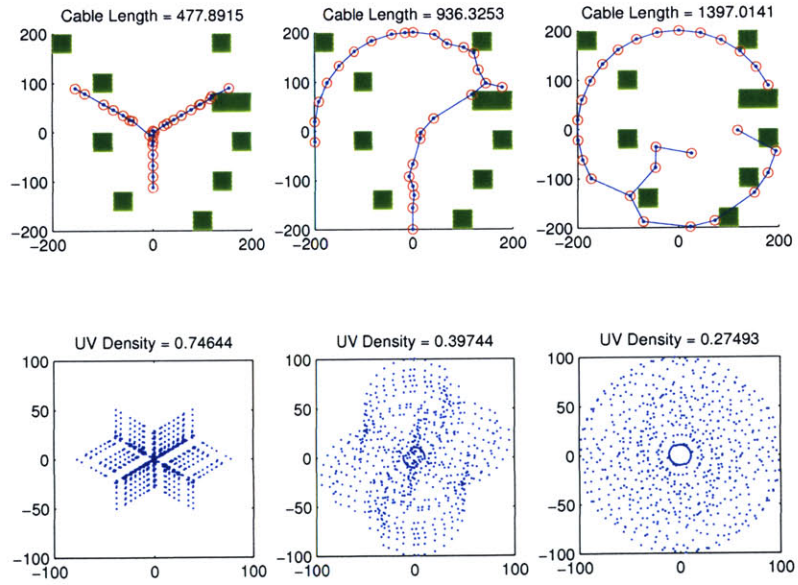


Figure 4-8: 27-station, 40km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

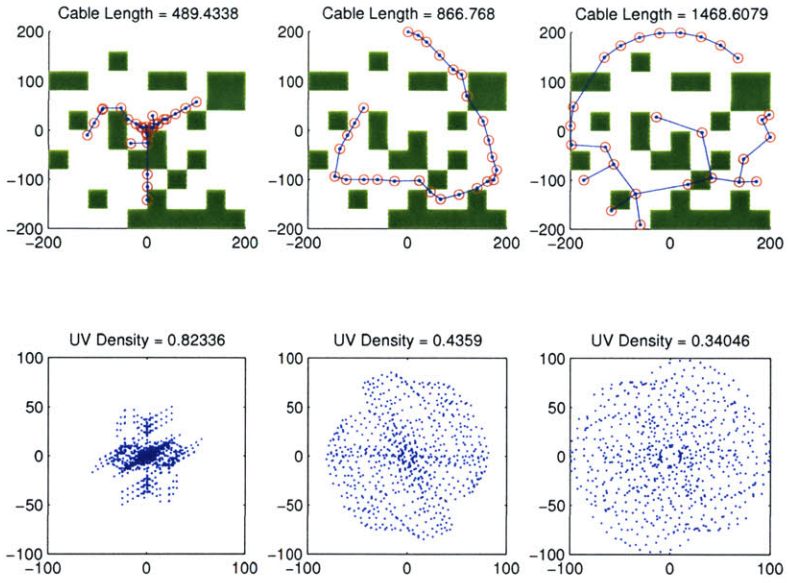


Figure 4-9: 27-station, 40km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

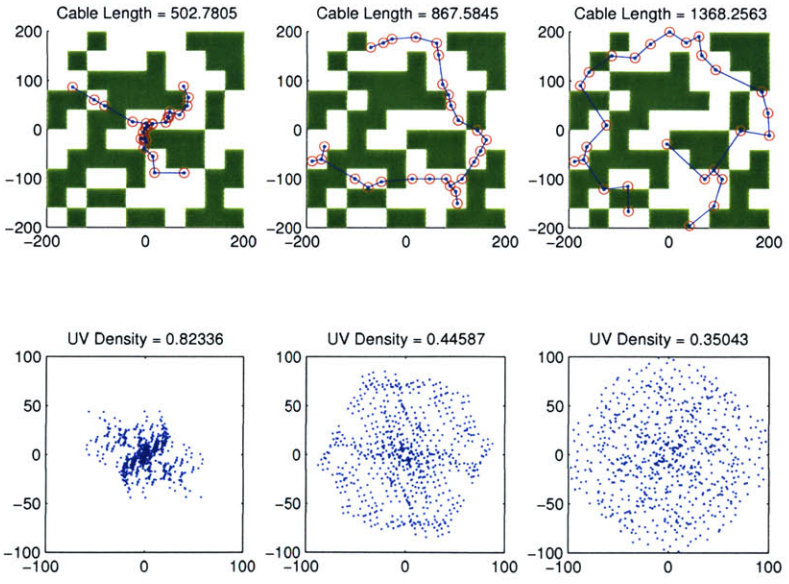


Figure 4-10: 27-station, 40km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

4.2.2 60-station Configurations

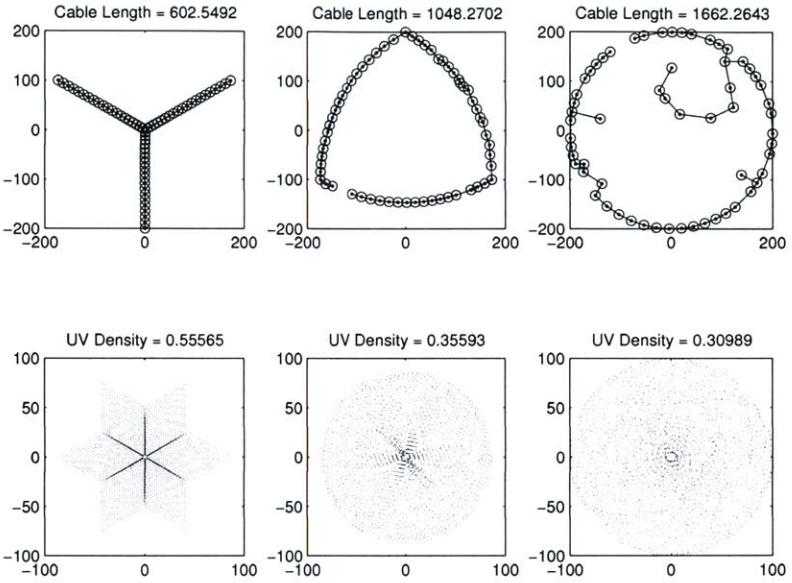


Figure 4-11: Unconstrained 60-station configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

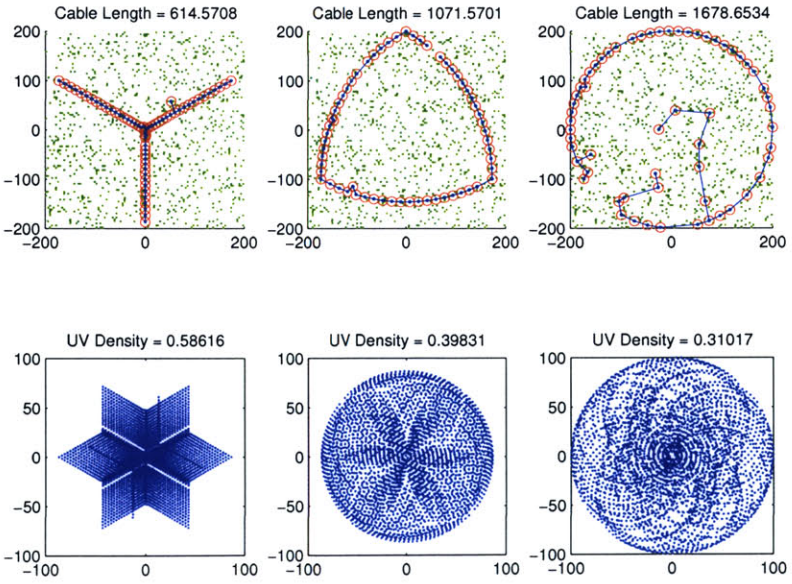


Figure 4-12: 60-station, 4km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

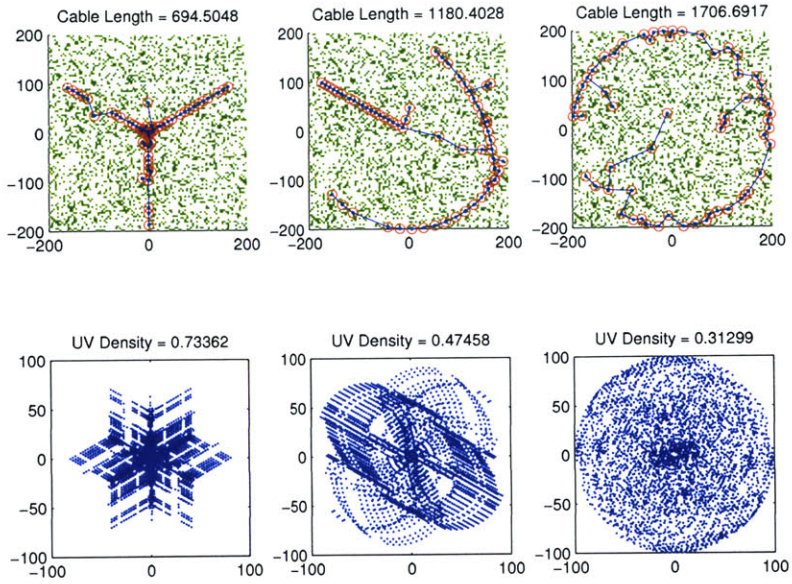


Figure 4-13: 60-station, 4km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

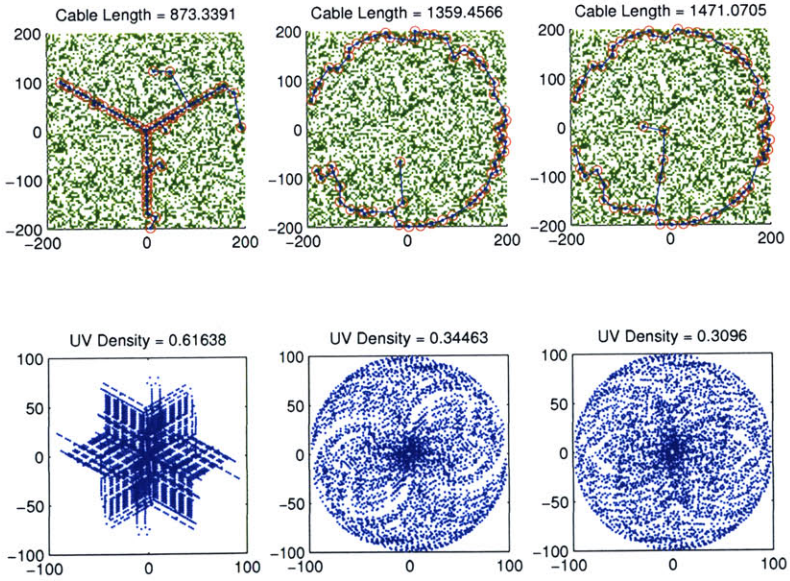


Figure 4-14: 60-station, 4km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

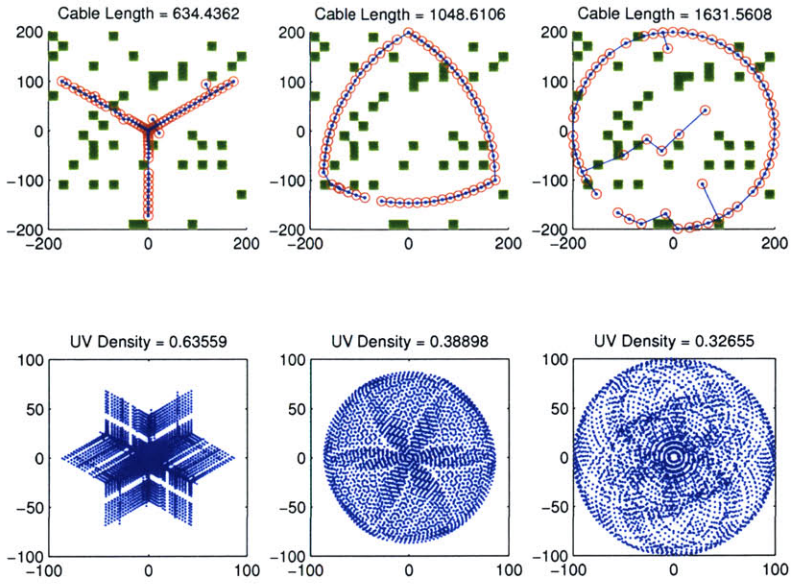


Figure 4-15: 60-station, 20km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

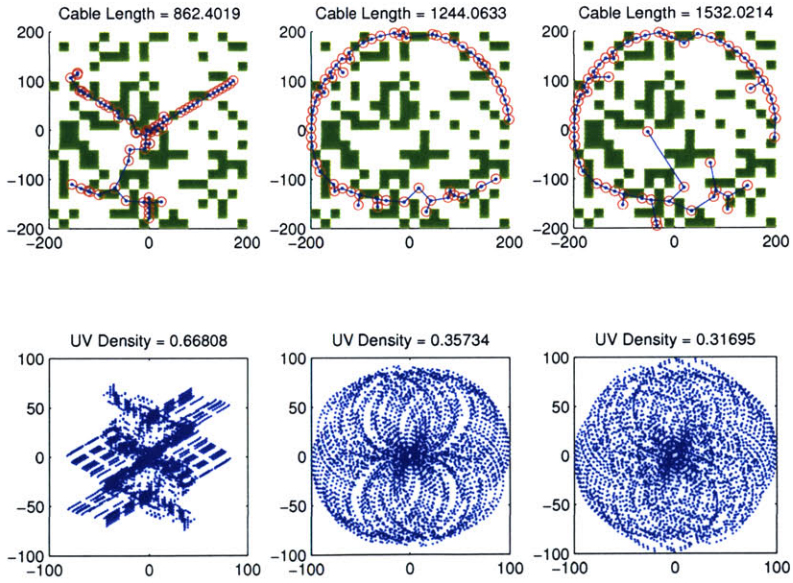


Figure 4-16: 60-station, 20km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

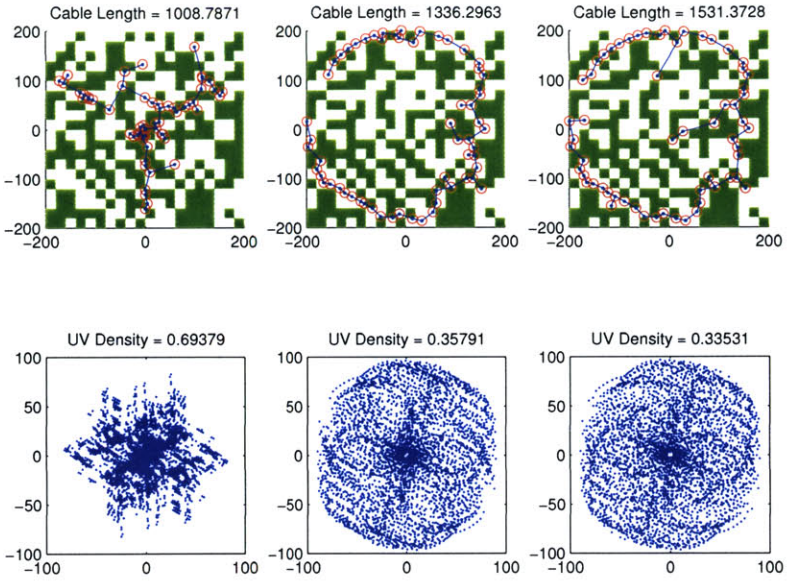


Figure 4-17: 60-station, 20km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

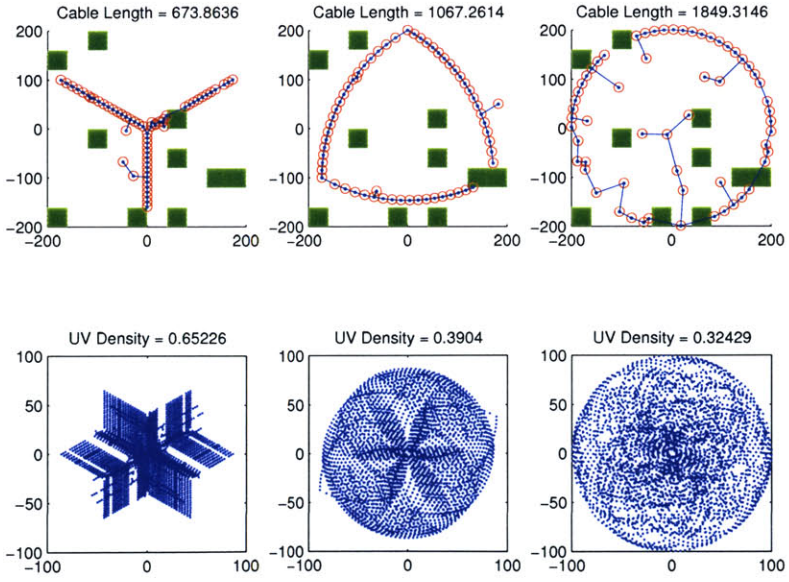


Figure 4-18: 60-station, 40km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

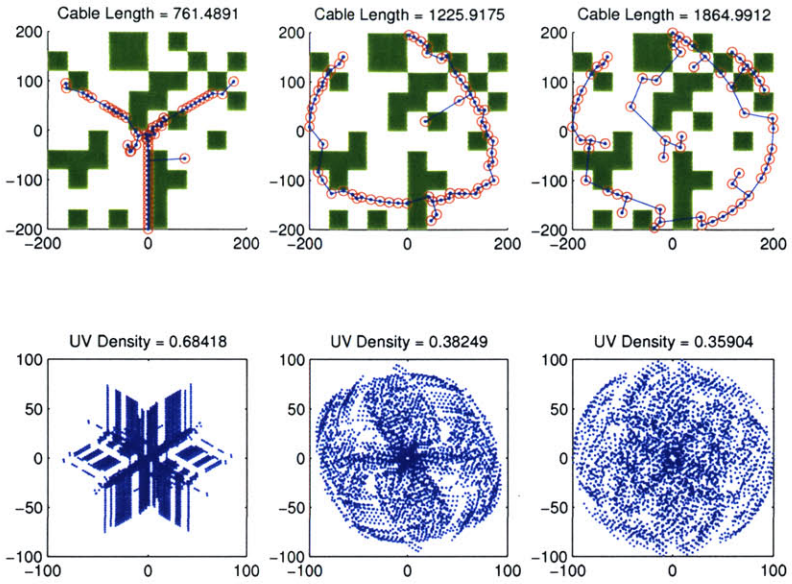


Figure 4-19: 60-station, 40km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

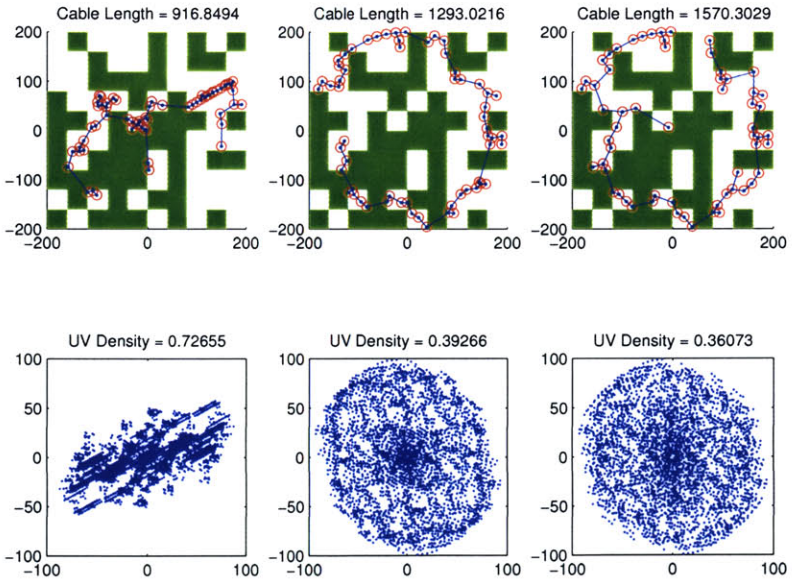


Figure 4-20: 60-station, 40km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

4.2.3 100-station Configurations

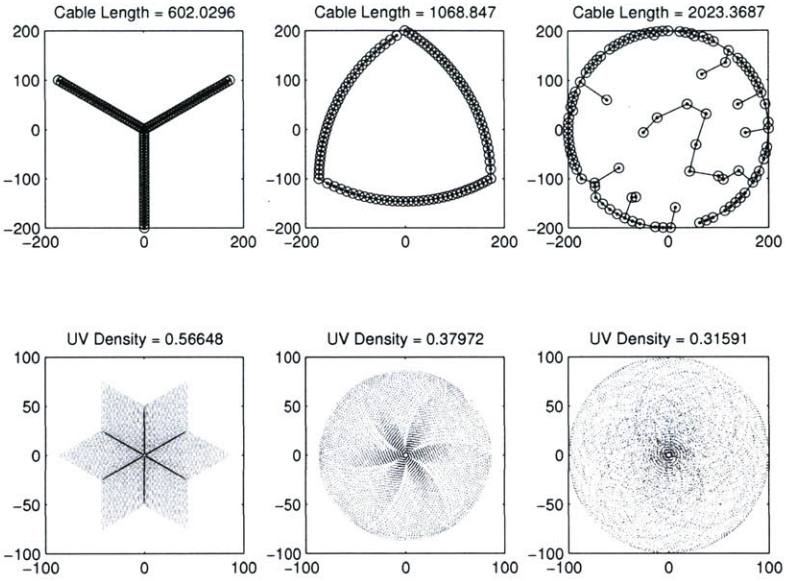


Figure 4-21: Unconstrained 100-station configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

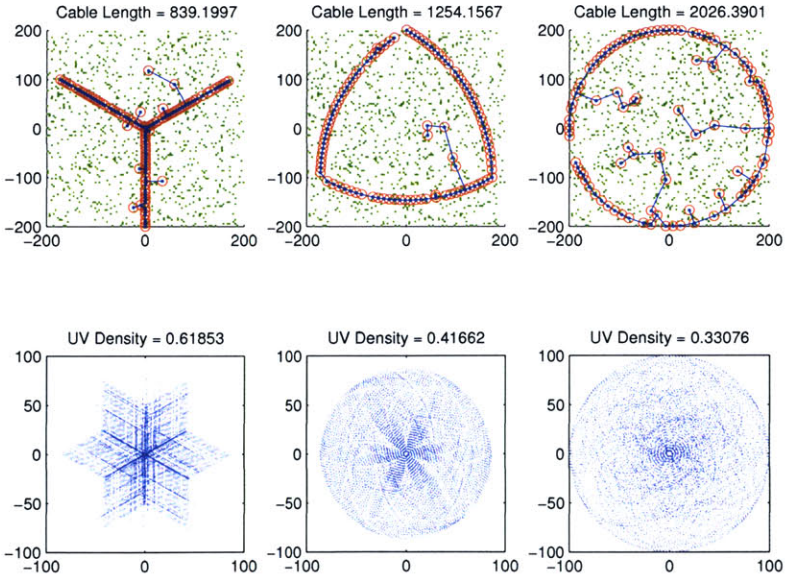


Figure 4-22: 100-station, 4km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

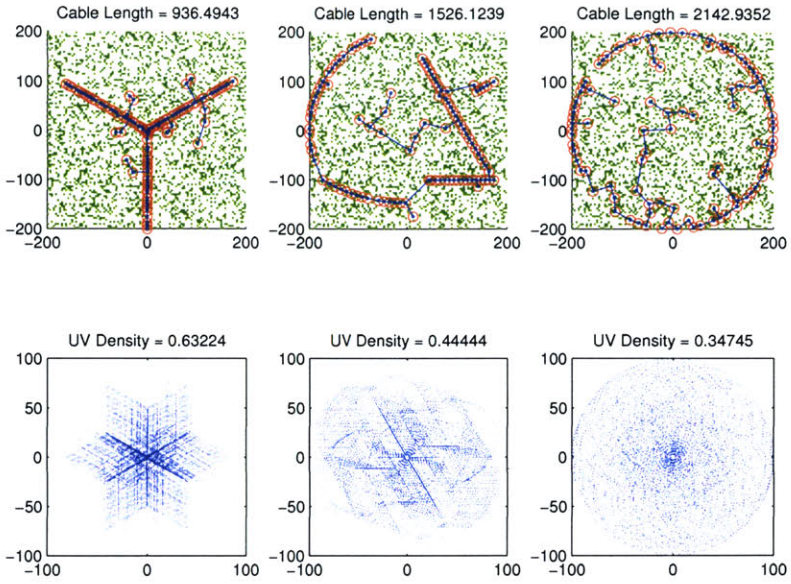


Figure 4-23: 100-station, 4km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

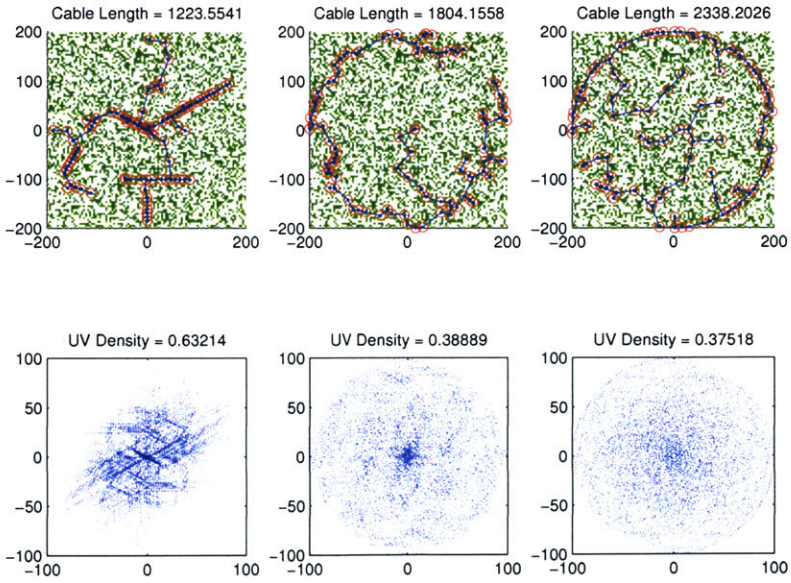


Figure 4-24: 100-station, 4km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

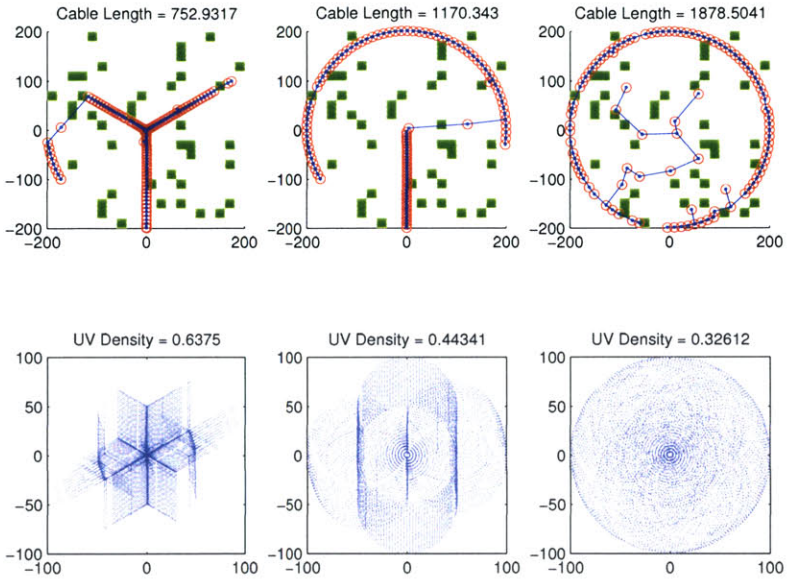


Figure 4-25: 100-station, 20km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

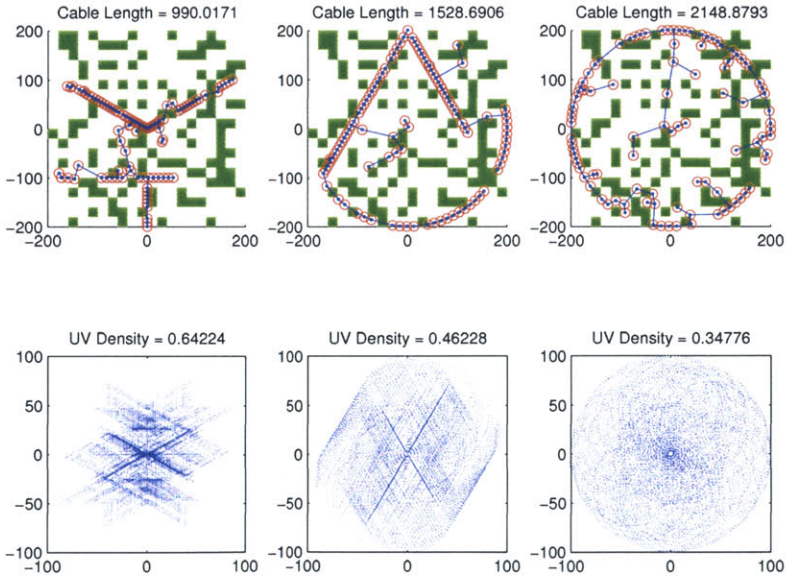


Figure 4-26: 100-station, 20km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

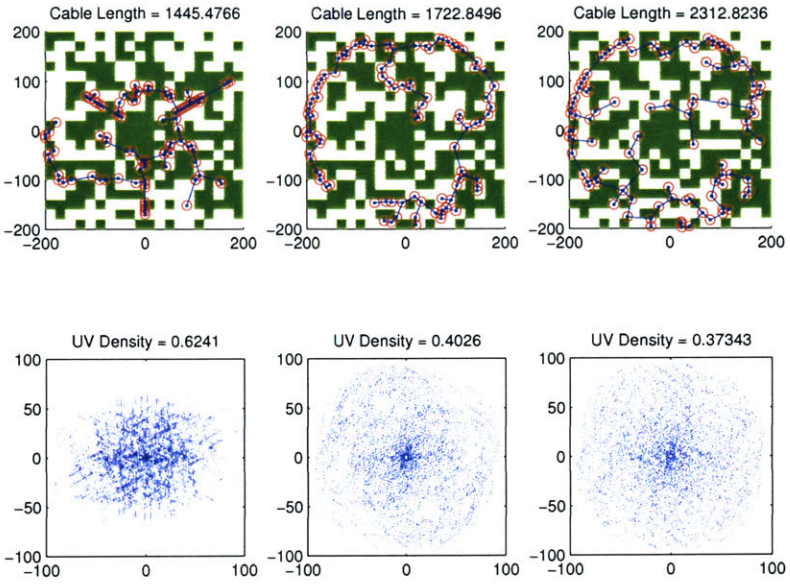


Figure 4-27: 100-station, 20km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

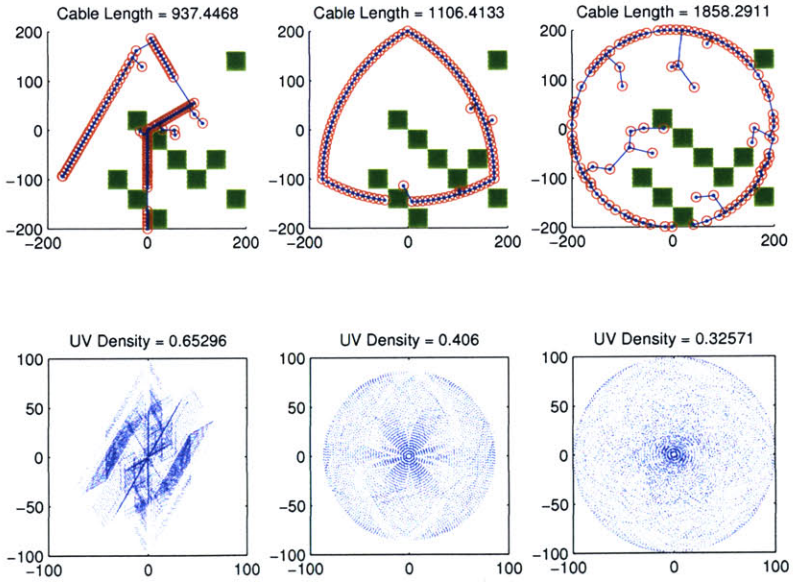


Figure 4-28: 100-station, 40km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

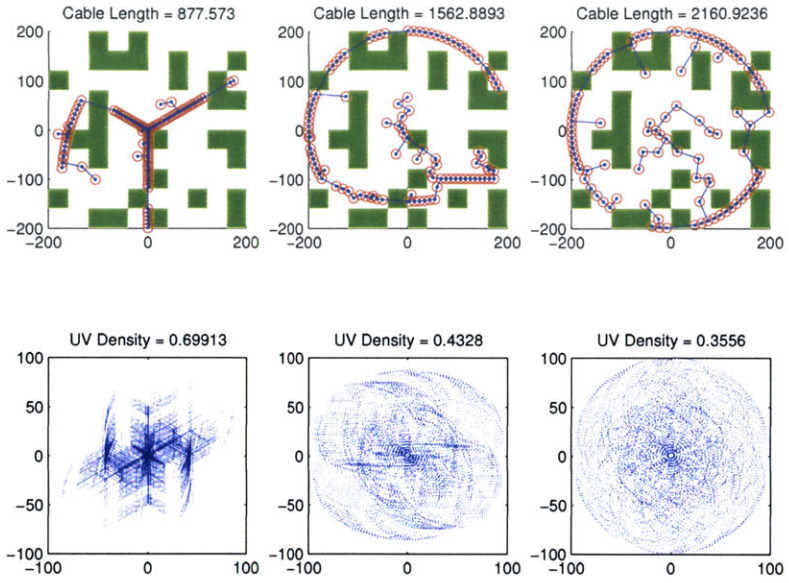


Figure 4-29: 100-station, 40km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

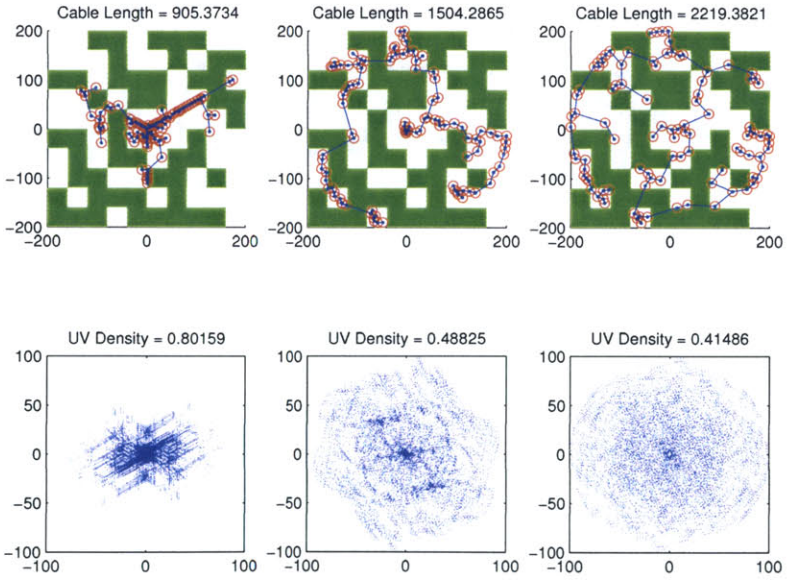


Figure 4-30: 100-station, 40km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

4.2.4 160-station Configurations

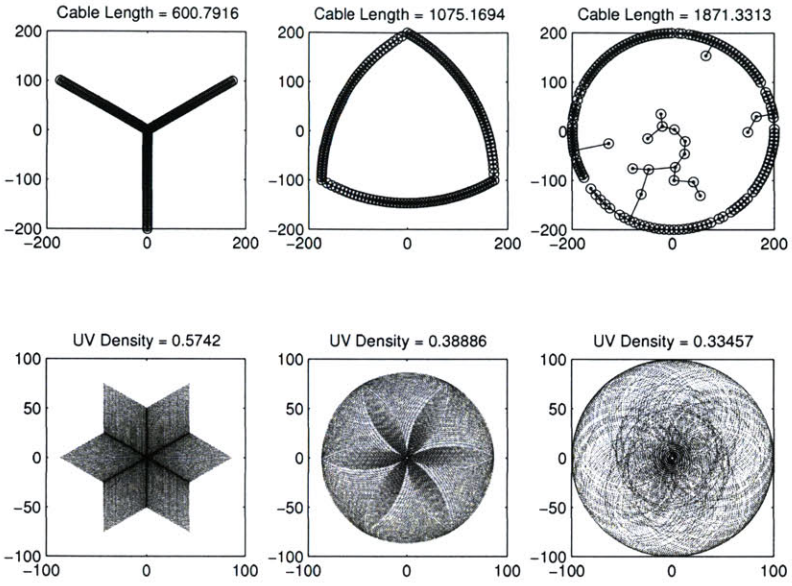


Figure 4-31: Unconstrained 160-station configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

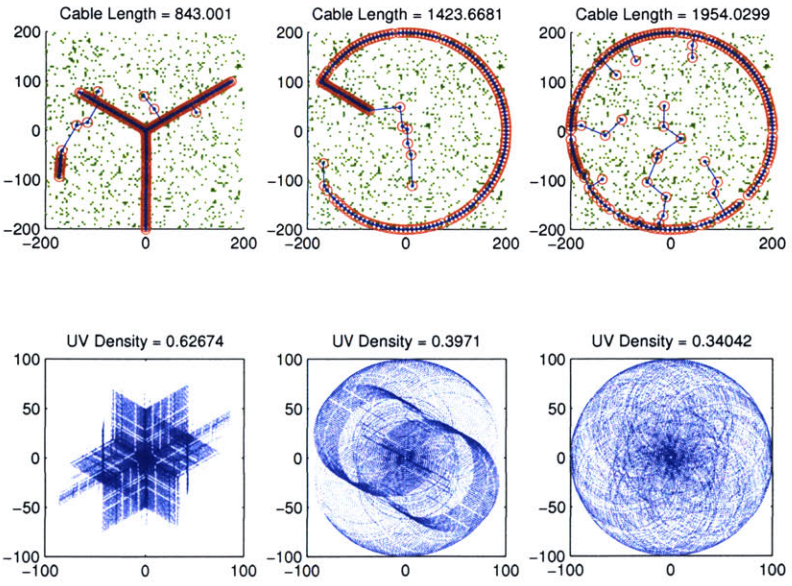


Figure 4-32: 160-station, 4km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

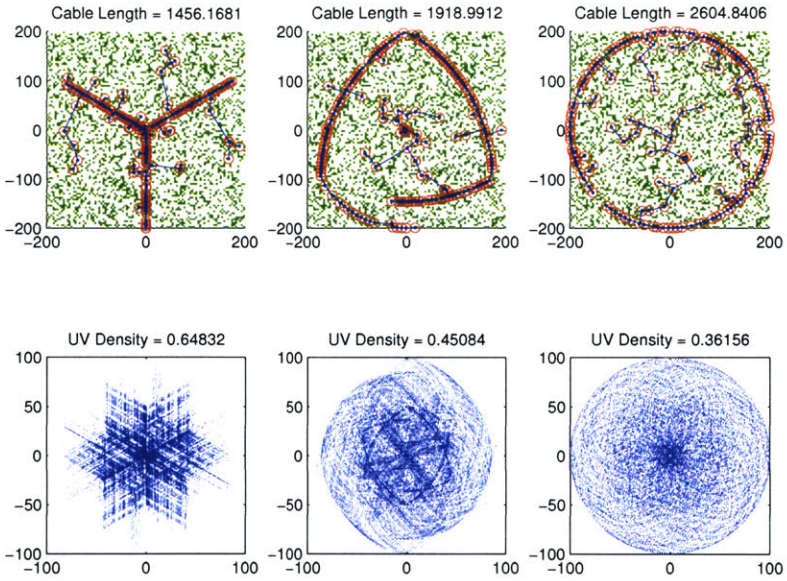


Figure 4-33: 160-station, 4km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

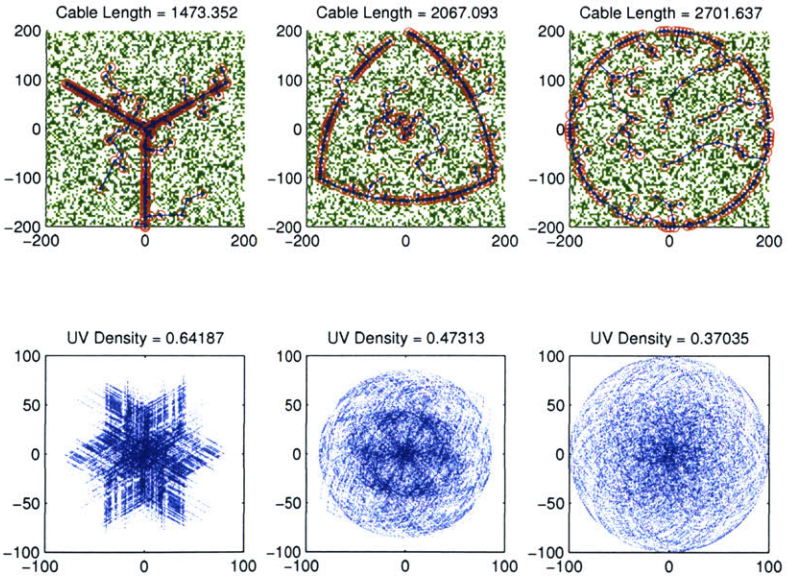


Figure 4-34: 160-station, 4km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

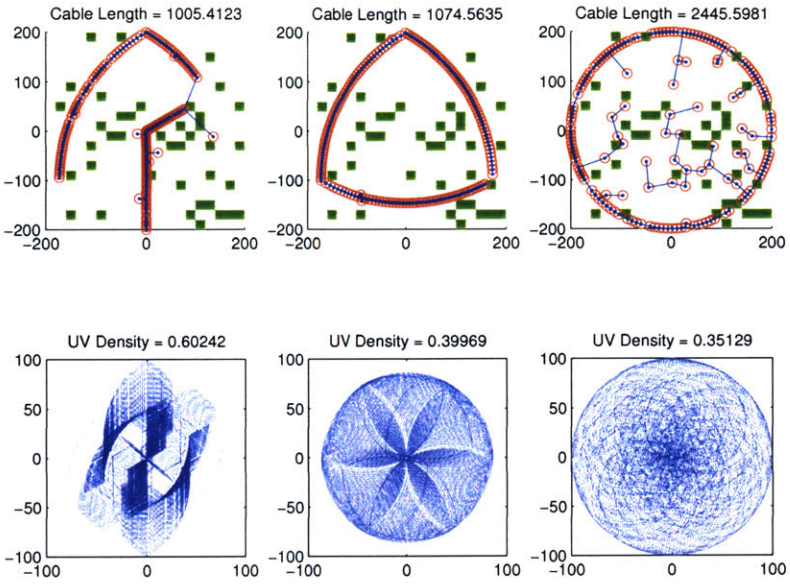


Figure 4-35: 160-station, 20km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

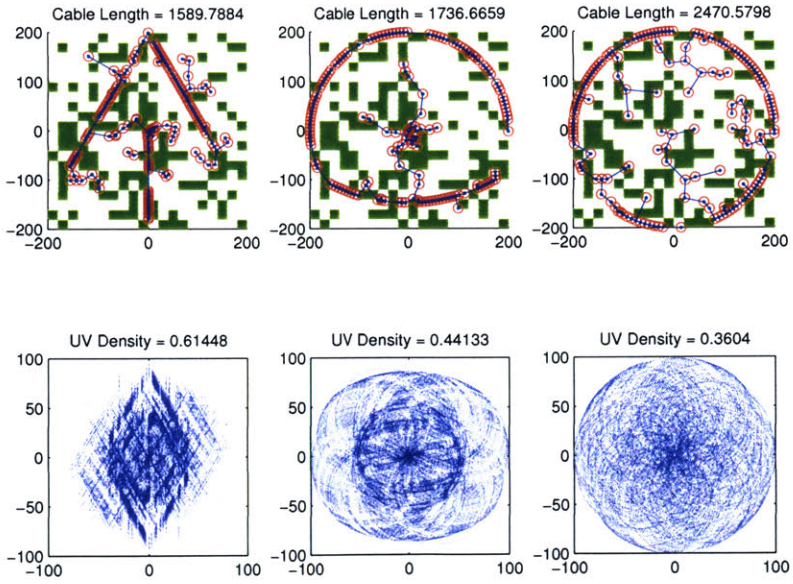


Figure 4-36: 160-station, 20km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

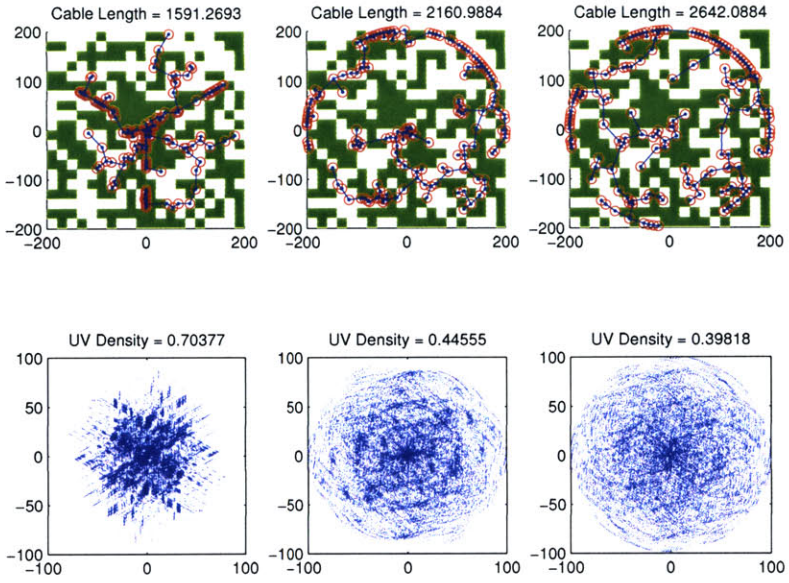


Figure 4-37: 160-station, 20km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

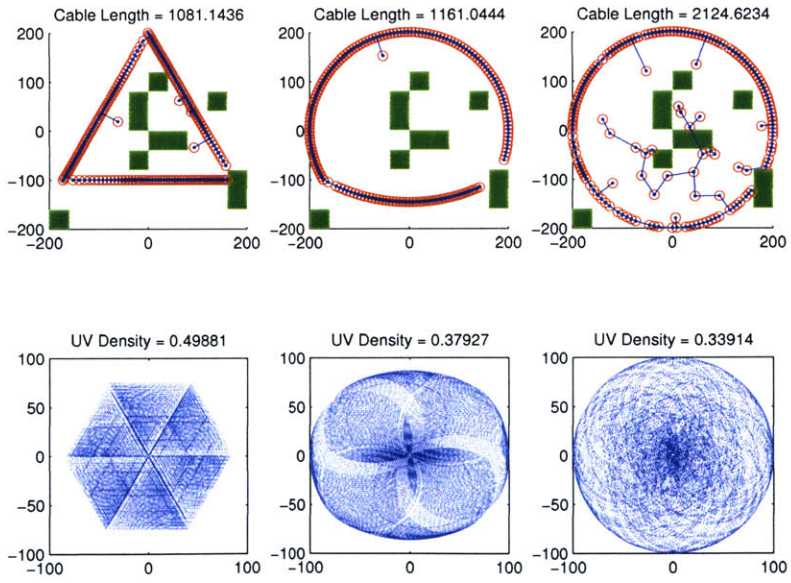


Figure 4-38: 160-station, 40km gridstep, 10% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

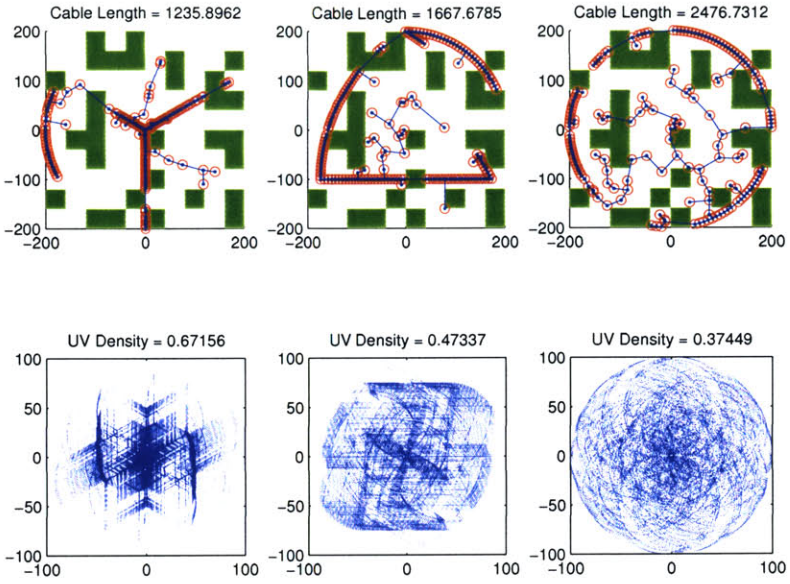


Figure 4-39: 160-station, 40km gridstep, 30% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

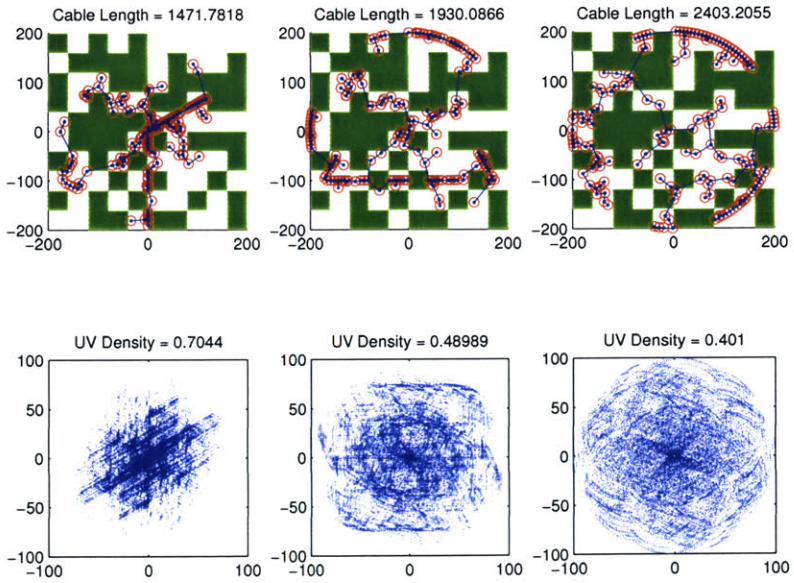


Figure 4-40: 160-station, 40km gridstep, 50% badzones (in green) configurations (top) with corresponding uv coverage (bottom). Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

4.3 Chapter 4 Summary

In this chapter the configurations are shown for the unconstrained configurations as well as the site constrained configurations. Unconstrained simulations were run to full convergence while site constrained simulations were all run for the same number of generations and population size. This was done to show comparisons specific to topics in Chapter 5. The constrained solutions were run with varying badzone sizes and percentages. The discussion of these effects on the objective spaces is discussed in Chapter 5.

Chapter 5

Convergence & Objective Spaces

5.1 Convergence Information

Since genetic algorithms are a heuristic form of optimization, it is never certain when absolute convergence is reached. Criteria can be set up to terminate the simulations after a major change in the design objectives has not occurred for a certain number of generations. This may not always be good termination criteria though, e.g. Figure A-2 shows that for nearly 1000 generations (between generations 1000 and 2000) there was very little change in either design objective until a mutation or favorable crossover created a significant change in the population which affected the cable length metric. The question becomes "what is a good termination criterion"? The termination criteria used in these simulations depends only on the predefined number of generations specified by the user, *ngen*. This was chosen for controlling the time to complete a simulation. Individual simulation convergence information can be found in Appendix A.3. Sections 5.1.1 & 5.1.2 will make comparisons in the convergence information with respect to the parameter $N_{stations}$ as well as site constraints.

Another method is possible if the user has a particular goal in mind. In most engineering problems there are specifications that need to be met and there is not much reward for creating designs that exceed the set specifications. If the user quantitatively knows what values of the design objectives are desired, the simulation can be terminated when all goals are met. This is known as goal programming.

5.1.1 Convergence as a Function of $N_{stations}$

The convergence plots can also bring some insight into the problem of destructive mating and not being able to advance the Pareto front past the initial seeds. Progressing from 27- to 160-stations in Figure 5-1 it can be seen that improvement from the initial seeds to the final solutions diminishes and even reverses for the 100- and 160-station simulations. It is also very noticeable that as the number of stations increases, the final population average for cable length gets markedly worse. Convergence is still attained, but at a much higher value than what was contained in the initial seeds. This lends credence to the hypothesis that destructive mating occurs between population members which breaks down the advancement of the Pareto front near the nadir-utopia point. It should be noted that anchor solutions are still expanded. This is a very interesting puzzle which arises for any highly nonlinear optimization problems. Section 6.3 will address this issue a little more. Further discussion of comparisons that can and cannot be made from convergence information will be discussed in Section 5.1.3 and pertain to issues dealing with varying $N_{stations}$ but also address deeper issues of convergence on a whole.

5.1.2 Effects of Site Constraints on Convergence

An interesting question is how the calculations required for computing site constraints affect the overall performance of the optimization process: is it worth computing and can it be computed without much computational cost? Section 5.2 will attempt to answer the first question, and hopefully on the basis of this entire thesis it will be shown that it is a worthwhile venture. In this section the computational feasibility of doing site constraint calculations will be shown.

Figures 5-2 - 5-5 show the convergence plots for all runs, unconstrained and constrained, sorted by the number of stations. Though individual simulation runs may be hard to discern, it can be seen that large scale convergence is achieved in all cases. Large scale convergence refers to the asymptotic behavior demonstrated by all cases. An interesting idea is that the initial seeds that the human operator places into the

framework is optimizing on the large scale, and the computer does the small scale fine-tuning which improves upon those large scale general designs. This will be discussed more in Sections 5.3 & 6.2. Even though full convergence has not yet been achieved, steady state evolution is suggested. Individual convergence plots are located in the appendix in Section A.3.

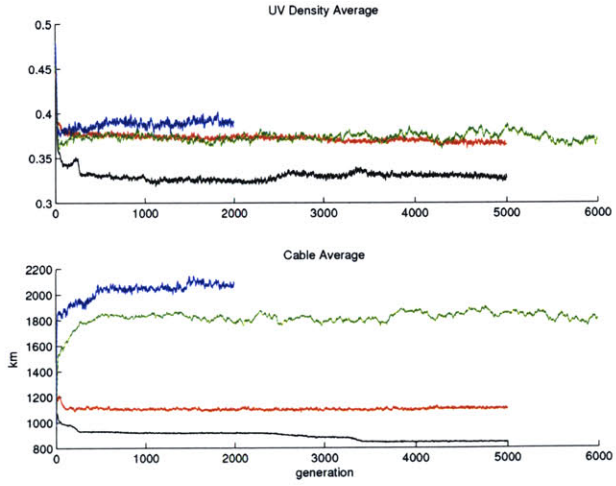


Figure 5-1: Unconstrained convergence information for 27-, 60-, 100-, and 160-stations. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation. 27-stations is denoted by the color black, 60-stations by red, 100-stations by green, and 160-stations by blue.

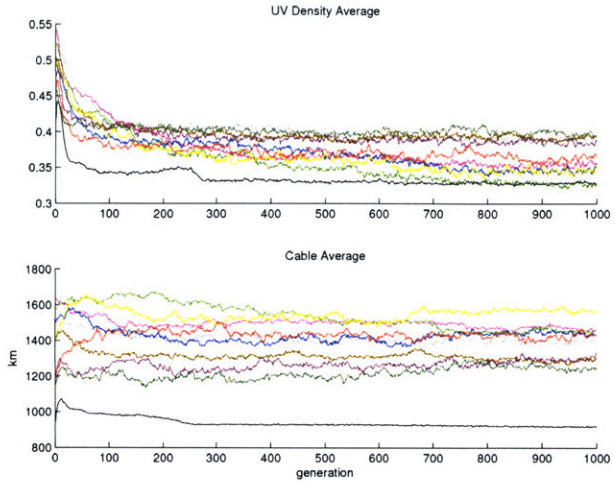


Figure 5-2: 27-station convergence information for all runs. Black indicates unconstrained, red indicates 4km *gridstep* with 10% badzones, yellow indicates 4km *gridstep* with 30% badzones, light green indicates 4km *gridstep* with 50% badzones, dark green indicates 20km *gridstep* with 10% badzones, blue indicates 20km *gridstep* with 30% badzones, magenta indicates 20km *gridstep* with 50% badzones, purple indicates 40km *gridstep* with 10% badzones, brown indicates 40km *gridstep* with 30% badzones, and gray indicates 40km *gridstep* with 50% badzones,

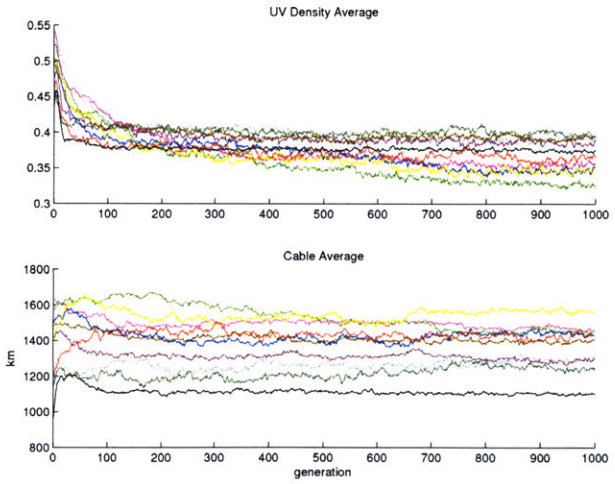


Figure 5-3: 60-station convergence information for all runs. Black indicates unconstrained, red indicates 4km *gridstep* with 10% badzones, yellow indicates 4km *gridstep* with 30% badzones, light green indicates 4km *gridstep* with 50% badzones, dark green indicates 20km *gridstep* with 10% badzones, blue indicates 20km *gridstep* with 30% badzones, magenta indicates 20km *gridstep* with 50% badzones, purple indicates 40km *gridstep* with 10% badzones, brown indicates 40km *gridstep* with 30% badzones, and gray indicates 40km *gridstep* with 50% badzones,

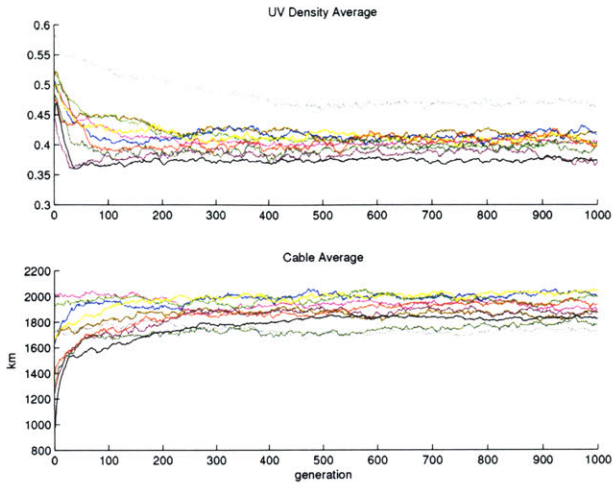


Figure 5-4: 100-station convergence information for all runs. Black indicates unconstrained, red indicates 4km *gridstep* with 10% badzones, yellow indicates 4km *gridstep* with 30% badzones, light green indicates 4km *gridstep* with 50% badzones, dark green indicates 20km *gridstep* with 10% badzones, blue indicates 20km *gridstep* with 30% badzones, magenta indicates 20km *gridstep* with 50% badzones, purple indicates 40km *gridstep* with 10% badzones, brown indicates 40km *gridstep* with 30% badzones, and gray indicates 40km *gridstep* with 50% badzones,

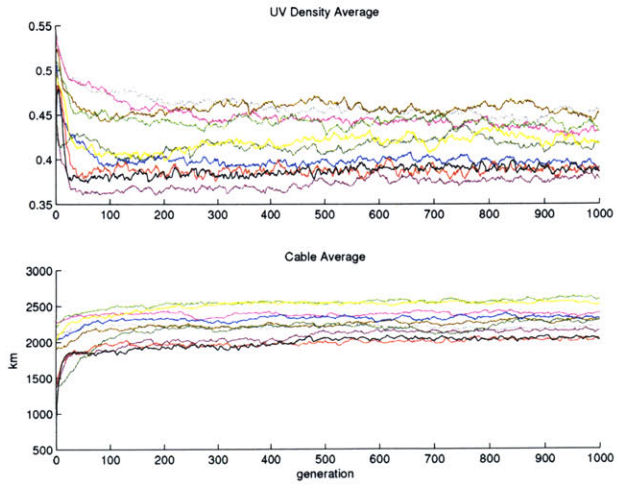


Figure 5-5: 160-station convergence information for all runs. Black indicates unconstrained, red indicates 4km *gridstep* with 10% badzones, yellow indicates 4km *gridstep* with 30% badzones, light green indicates 4km *gridstep* with 50% badzones, dark green indicates 20km *gridstep* with 10% badzones, blue indicates 20km *gridstep* with 30% badzones, magenta indicates 20km *gridstep* with 50% badzones, purple indicates 40km *gridstep* with 10% badzones, brown indicates 40km *gridstep* with 30% badzones, and gray indicates 40km *gridstep* with 50% badzones,

It should be noted that when looking at the convergence plots, a comparison cannot be made about how well the performance of the arrays changes with number of stations using the current metric. The performance metric, M, is a relative measure upon the number of stations. Thus the metric normalization changes as $N_{stations}$ changes. Intuitively it would make sense that as the station numbers increased, the performance of the array would increase. The metric depends on the uv density, and the number of uv points goes as N^2, so the overall density would thus increase by the same rate. What is interesting is if a configuration with a smaller number of stations can out perform a configuration with a larger number of stations. This would require a new way of comparing configurations and will be suggested in Section 6.1.

What can be compared from the convergence information is the cable length as a function of the number of stations. It may seem intuitively that the fewer number of stations, the lower the cable length, but highly geometric designs have stations fall on pathways that are shared by a large number of stations, as can easily be seen from the plot of initial seeds in Figure 2-1. The data in Figure 5-1 suggests that on average cable lengths are higher for a larger number of stations.

5.1.3 Convergence in Multiobjective Optimization

An interesting caveat arises from the fact that the convergence plots are made for a multiobjective optimization. What occurs in a multiobjective optimization? Configurations are constantly being changed based upon more than one objective. The convergence plots are made from taking the average of the metrics for the population for a given generation and plotted. In the average a lot of information is lost about what is actually happening in the optimization process. Things can be very stagnant, which is what we expect in the traditional single objective multiobjective problem, and thus the average represents the overall population very well. Things may be highly dynamic throughout the population, or one part of the population may be very static while other parts are still very dynamic, but in the end all that is seen is the population average. In all likelihood the asymptotic behavior shown in the convergence plots occurs because of stagnation in the evolution of the process, but

this is not certain from convergence plots alone. This leads to a possible study of how convergence should be defined in a multiobjective problem. Plotting the best solutions for all the metrics may be another way to determine convergence. These types of plots will show convergence in the anchor solutions. Figures 5-6 & 5-7 show how the convergence compares between using the average of the metrics of the population at a given generation versus using the best metric of the individual in a generation.

Further discussion of this topic will occur in Section 6.3, but it can be said here that convergence may not be as much of an issue as the evolution of the population. What has been done as a part of this study and may lend insight into a possible solution is that of tracking the evolution of the population through the objective space, which is done more affectively in the objective space plots. In essence this is another area of research that needs to be studied further.

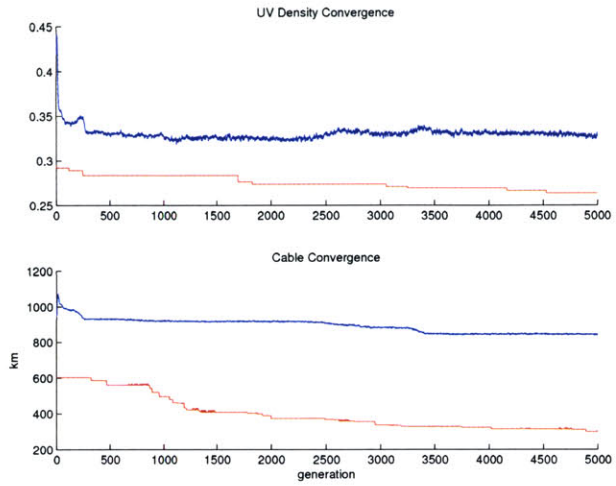


Figure 5-6: 27-station convergence information using best of population. The convergence of the population using the average over all members in a given generation is given in blue, while the convergence of the best of the population is given in red.

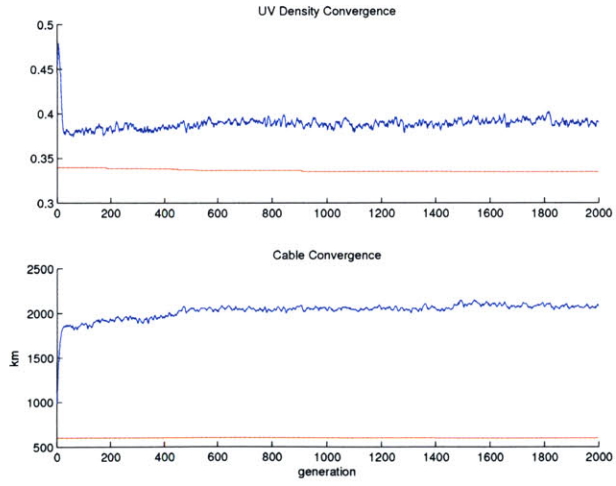


Figure 5-7: 160-station convergence information using best of population. The convergence of the population using the average over all members in a given generation is given in blue, while the convergence of the best of the population is given in red.

5.2 Objective Spaces

Figures 5-8 - 5-11 show the 27-, 60-, 100-, and 160-station objective spaces for the unconstrained simulations. Cable length is on the x-axis, while the uv density metric, M, is on the y-axis. UV density metric values can range from close to zero (filled aperture), to nearly one (very poor uv coverage). Since the best solutions presented here asymptote at above 0.25, this is given as a lower boundary on the axis. Since VLA-like designs were chosen as the minimum cable configurations (uv density values ≈ 0.55), the upper boundaries of the uv density metric were cut off at an arbitrary value of 0.75 to show a few designs which achieved lower cable lengths and lower performance but are still formally Pareto optimal. As discussed in an earlier section (Section 4.2) there needs to be a minimum performance that designers wish to consider and a cut off is chosen. In the design of constrained configurations, this was not necessarily true because the site constraints broke down the highly geometric designs initially to concur with the constraints. The minimum cable configurations were chosen at a value close to that of the VLA.

The objective spaces presented show a wide range of Pareto optimal designs. In all cases a concave Pareto front developed with decreasing marginal returns as designs improved in either metric. This leaves the choice up to designers for the type of array that can be afforded. If cable length is not an issue in the array design, then ring-like solutions with inward reaching arms are a good choice. If cable length is a major issue, then slightly randomized VLA-like configurations may be the best choice. If a trade-off is desired, the Reuleaux triangle configurations, or other hybrid designs near the nadir-utopia point may be chosen. Appendix A.2 shows an entire family of Pareto front solutions from the minimum cable anchor solution through to the best performance cable solution. As more objectives are added to the multiobjective design problem, different Pareto surfaces may give new tradeoffs between objectives, particularly if the array should be grown over time (see Section 6.3).

Interesting results have stemmed from studying and comparing the unconstrained simulations and how they differ with the number of stations. Section 5.2.1 discusses

the unconstrained simulations and what can be learned about the nature of the problem before constraints are introduced. Section 5.2.2 discusses how site constraints affect the optimization process.

5.2.1 Objective Spaces of Unconstrained Simulations

An interesting feature of the unconstrained objective spaces is that for arrays with a large number of stations, well-known highly geometric configurations are optimum and occupy particular regions on or near the Pareto front (non-geometric seeds and their comparison to geometric seeds will be discussed in Section 5.3). There is a progression from Y-configurations, to triangles and Reuleaux triangles, to circles. For the smaller number of stations, optimum solutions are found that are significantly better than the initial seeds. For the larger number of stations the array performance was improved, but it was difficult to find any designs that simultaneously improved array performance and shortened cable length; i.e., the Pareto front could not be advanced near the nadir-utopia point. It should be noted, however, that these conclusions are specific to the two metrics considered in this study. Introducing new design objectives may shift the optimum designs away from geometric arrays.

Perturbations from ideal geometries and reduction in unnecessary components of the initial seeds were sufficient to improve the designs. Why were there no new topologies found? The initial seeds into the population are highly geometric. Highly geometric arrays have smooth pathways for cable configurations to follow, thus already having quite a low cable length compared to very similar designs that have random perturbations from the ideal geometries and similar array performance. Also, in a sense there is an entropy that arises in doing optimizations with a large number of design variables. It is easy to go from highly geometric shapes to more randomized shapes, but statistically it is difficult to create highly geometric shapes from more randomized shapes during crossover and mutation of station configurations. Highly nonlinear objective spaces, such as this one, also pose the problem that if one good array configuration swaps station placement information with another good array configuration, the resulting array is not necessarily going to be an improvement, but

more than likely it will be less optimum if the arrays are very different from each other. We will refer to this as destructive mating. These problems are difficult to address because they seem to suggest that expensive Monte Carlo exploration of the objective space should be carried out. Perhaps greater improvement can come from implementing different genetic algorithm techniques. An example is selective mating [10], a modification to the mating algorithm which restricts designs which are too dissimilar from exchanging information. Further ideas to solve these problems will be discussed in Future Work (Section 6.3).

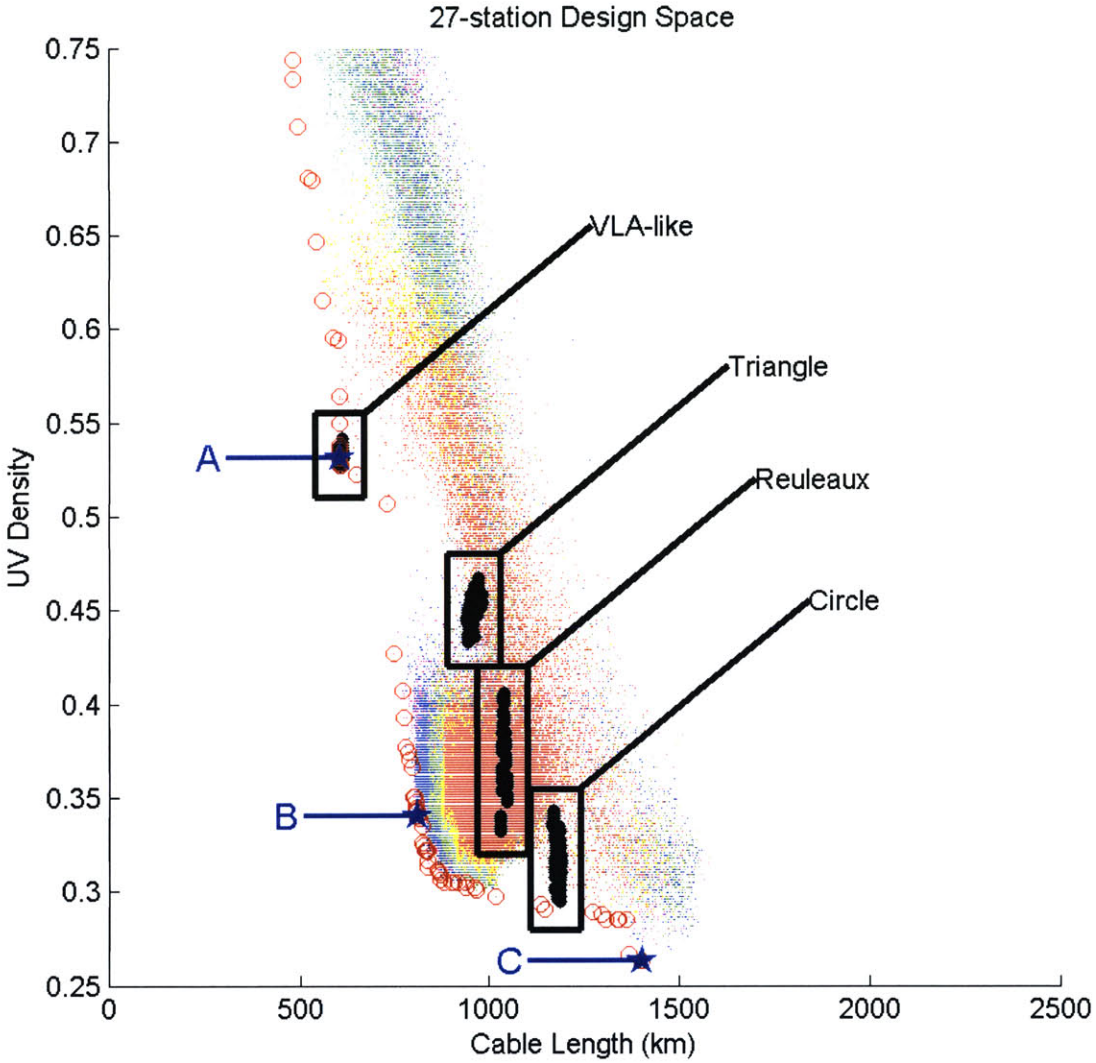


Figure 5-8: Unconstrained 27-station objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles. The blue stars are the (**A**) minimum cable, (**B**) nadir-utopia, and (**C**) maximum performance configurations which are shown in Figure 4-1.

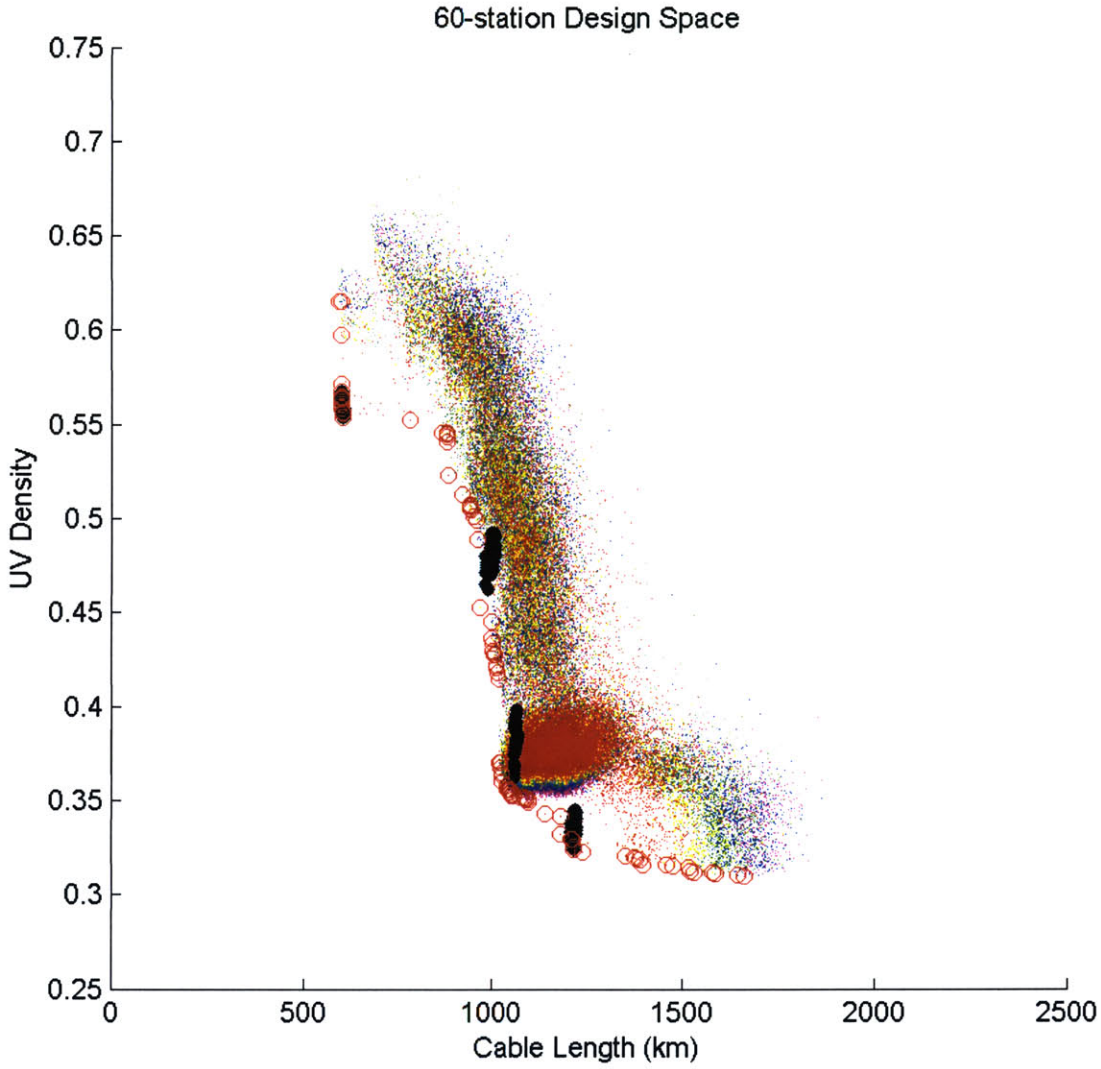


Figure 5-9: Unconstrained 60-station objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

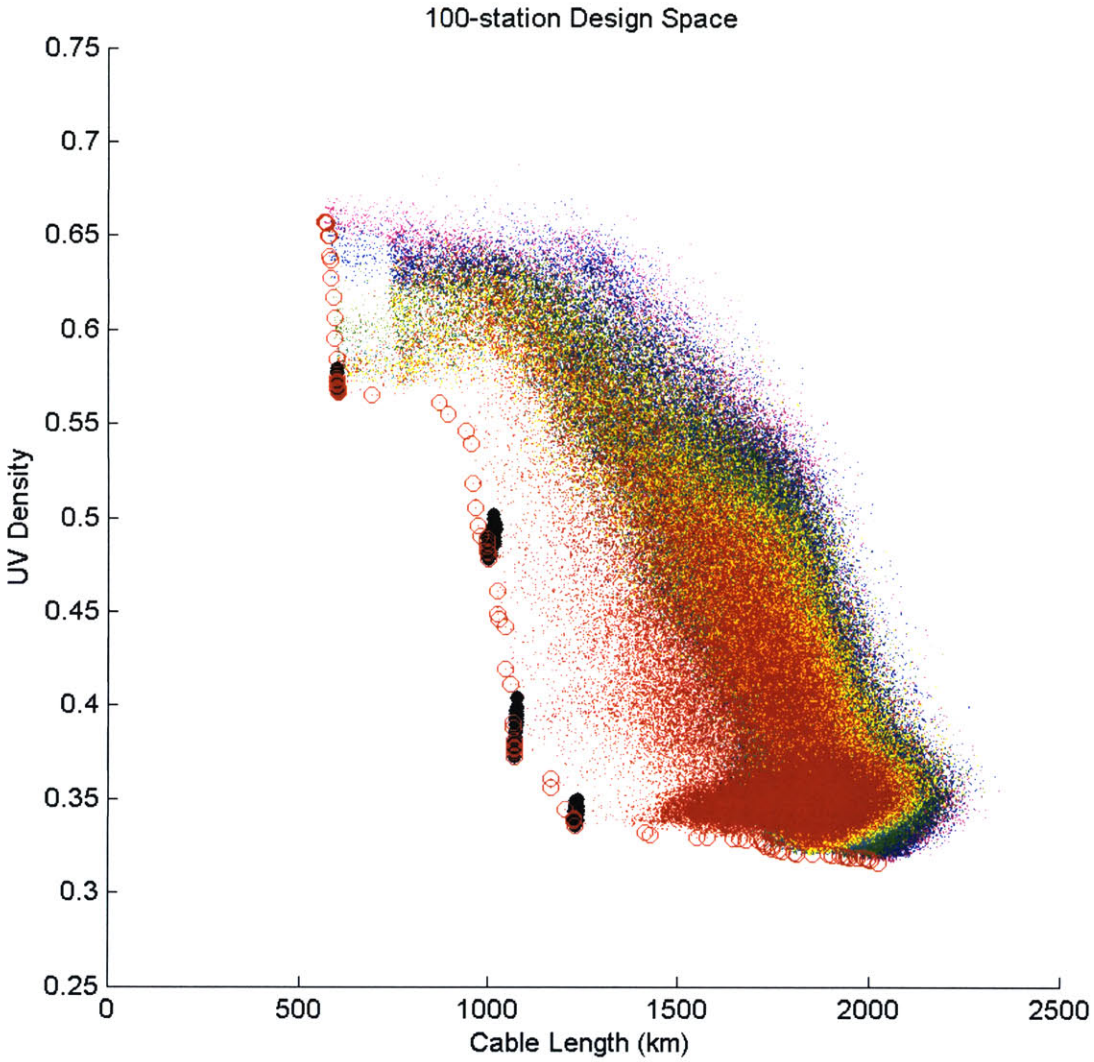


Figure 5-10: Unconstrained 100-station objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

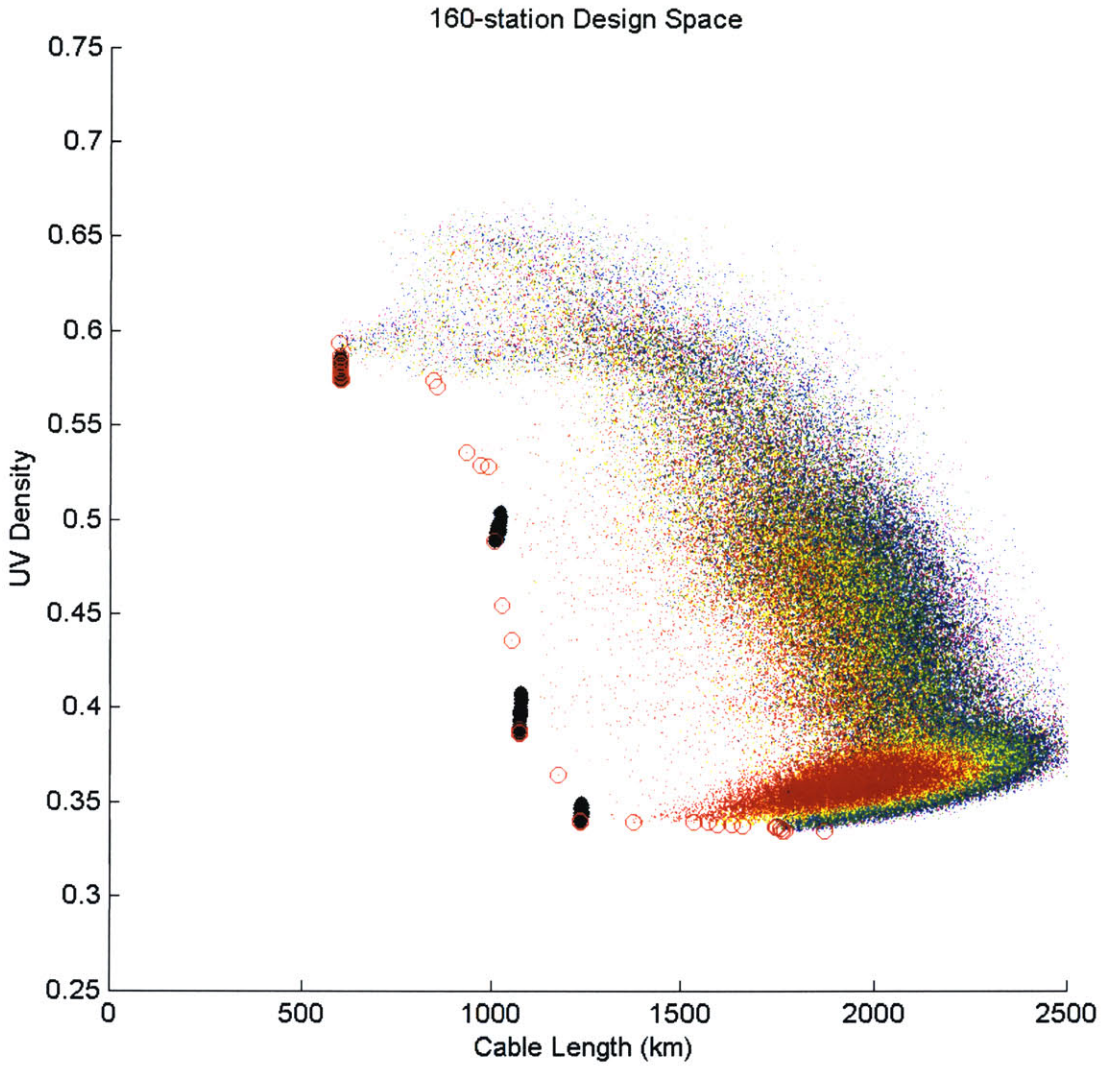


Figure 5-11: Unconstrained 160-station objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

5.2.2 Effects of Site Constraints on Objective Space

There are three interesting parameters to consider when assessing the effects of site constraints on the objective spaces. One is the parameter *percent*, which determines the percentage of badzones. Another is the parameter *gridstep*, which determines the size of the badzones. The third interesting parameter to consider is the number of stations, $N_{stations}$, and how adding more stations affects the design objectives along the Pareto front for both the *percent* and *gridstep* parameters.

It should be noted that the site constrained simulations were not always run to full convergence. All site constrained simulations were run for the same number of generations and population size for convergence comparison studies as well as computational considerations, but as can be shown from studies done in Section 5.1, it was shown that convergence of most simulations were comparable enough to demonstrate the affects of site constraints on the objective spaces without having to achieve full convergence.

Effect of *Percent* Parameter

Figures 5-12 - 5-23 show plots of the Pareto fronts while varying the site parameter *percent*. Plots were made for different *gridstep* sizes and number of stations. As can be seen from the plots, as the number of stations increases, the *percent* parameter pushes back the Pareto fronts more and more in both metrics. Conceptually it can thus be said that as the percentage of badzones increases the performance and cable metric are affected negatively, and to a greater extent as more stations are trying to occupy the same limited space. Intuitively this makes sense since the constraints are becoming more and more restrictive.

Both metrics are degraded, but it is obvious to see that the cable length metric is affected more than the performance metric. With more and more site constraints being imposed, it is harder to place stations as efficiently as can be done in an unconstrained simulation that can house highly geometric designs. Good performance can still be achieved by utilizing the given space as much as possible.

It should be noted that the unconstrained cases were run for more generations and with a higher population number, so they do not coincide with site constraint simulations necessarily. They have been shown as a reference case to demonstrate evolution properties as a function of *percent*.

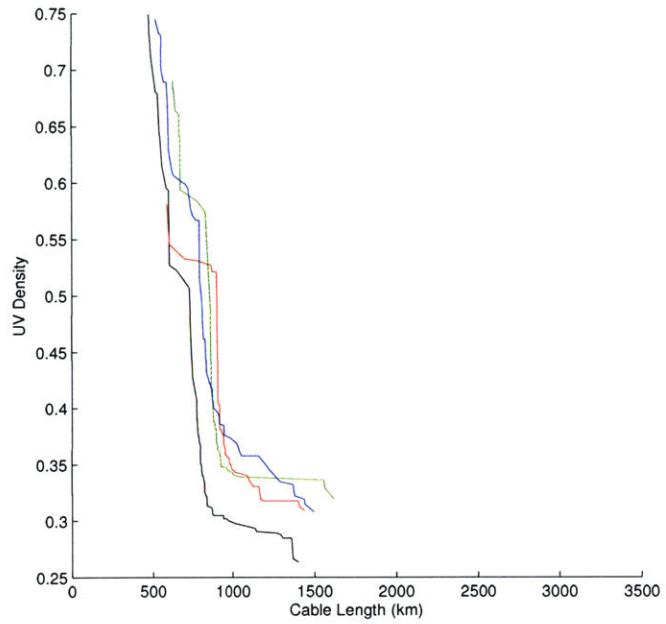


Figure 5-12: 27-station objective space varying *percent*, 4km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

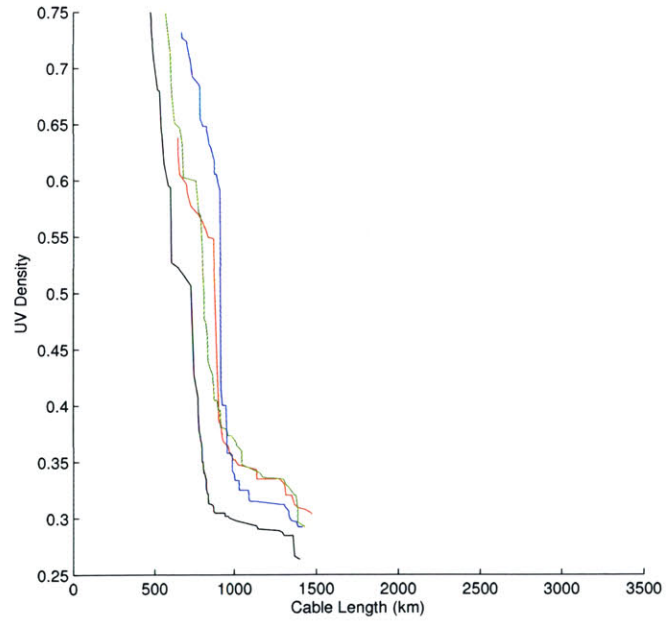


Figure 5-13: 27-station objective space varying *percent*, 20km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

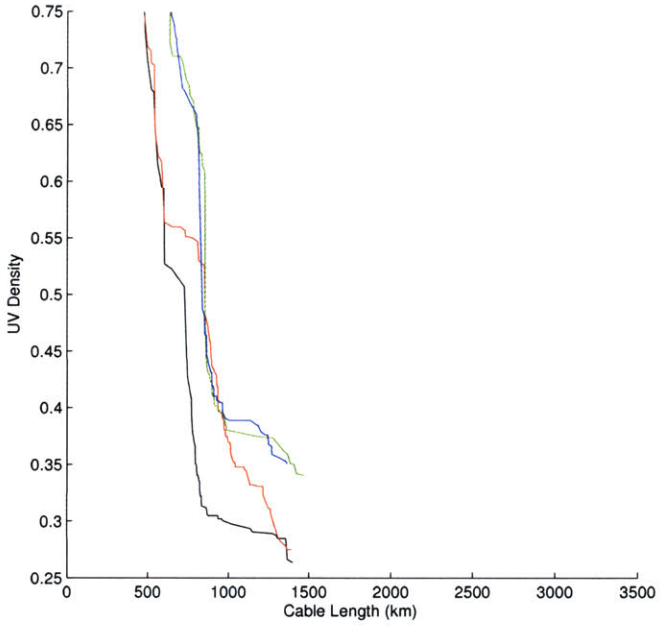


Figure 5-14: 27-station objective space varying *percent*, 40km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

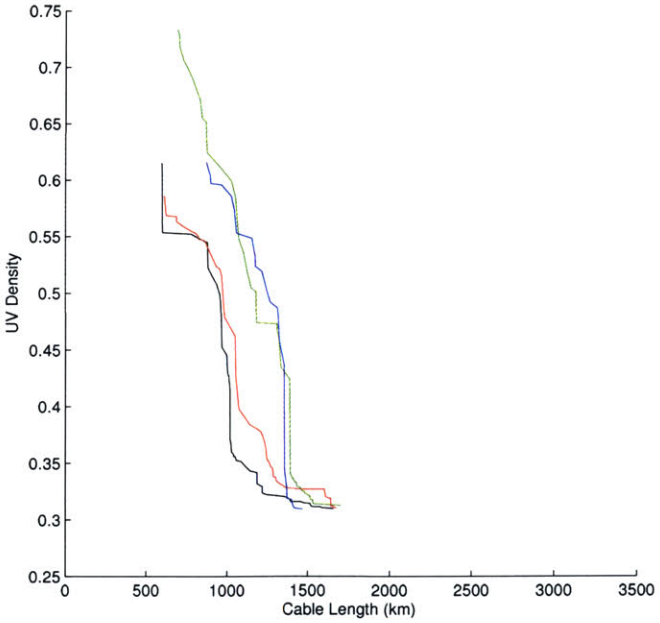


Figure 5-15: 60-station objective space varying *percent*, 4km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

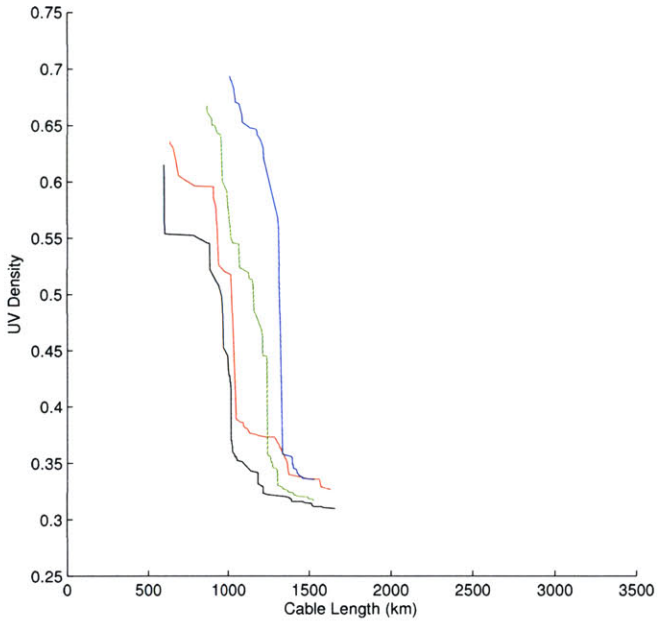


Figure 5-16: 60-station objective space varying *percent*, 20km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

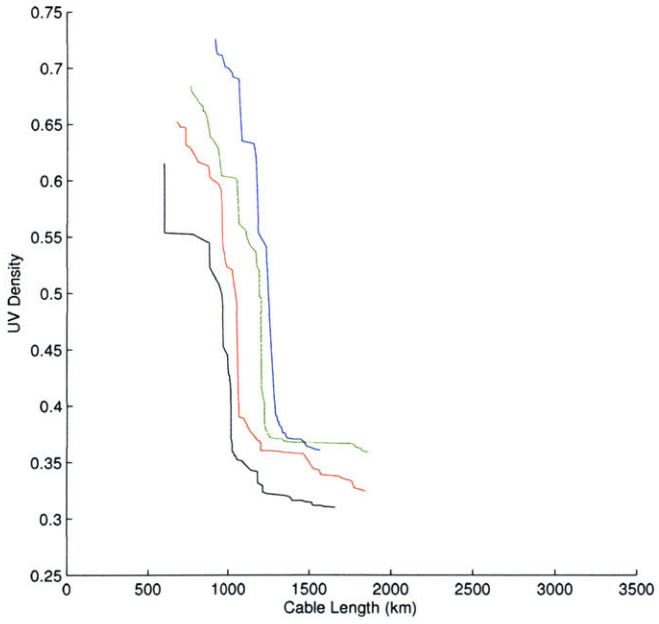


Figure 5-17: 60-station objective space varying *percent*, 40km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

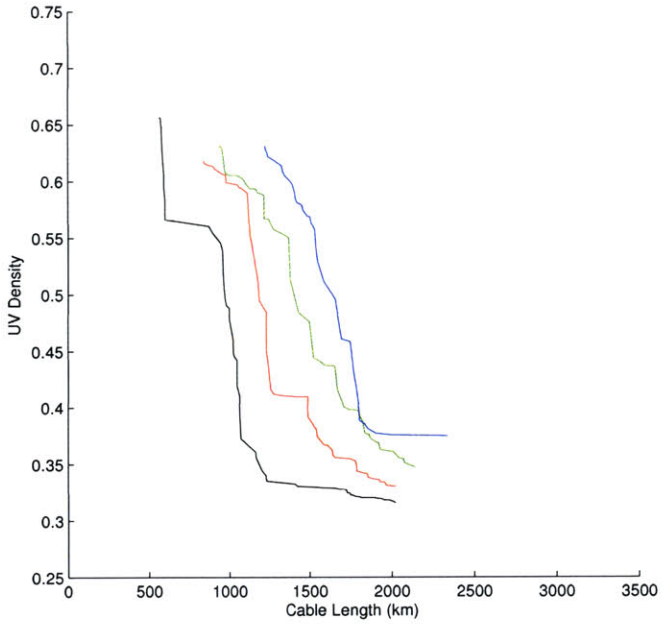


Figure 5-18: 100-station objective space varying *percent*, 4km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

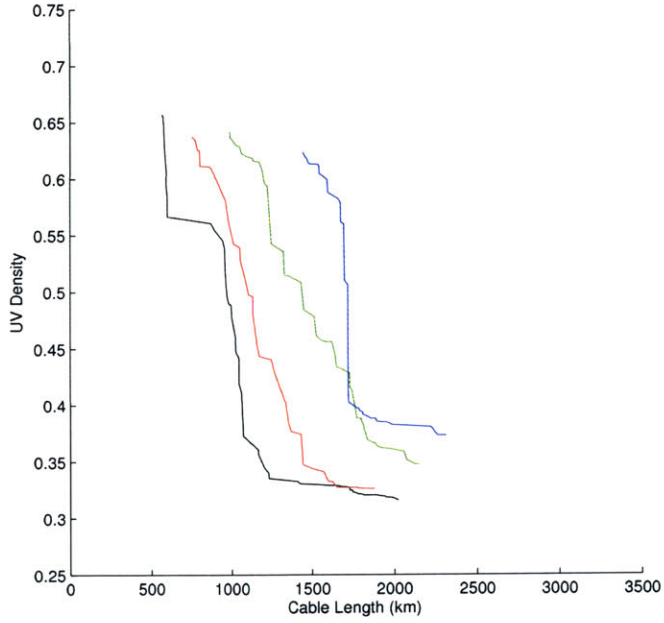


Figure 5-19: 100-station objective space varying *percent*, 20km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

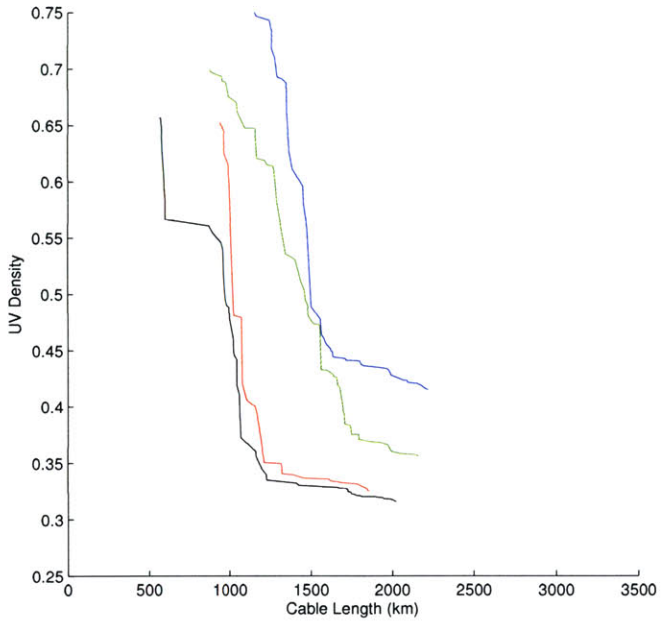


Figure 5-20: 100-station objective space varying *percent*, 40km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

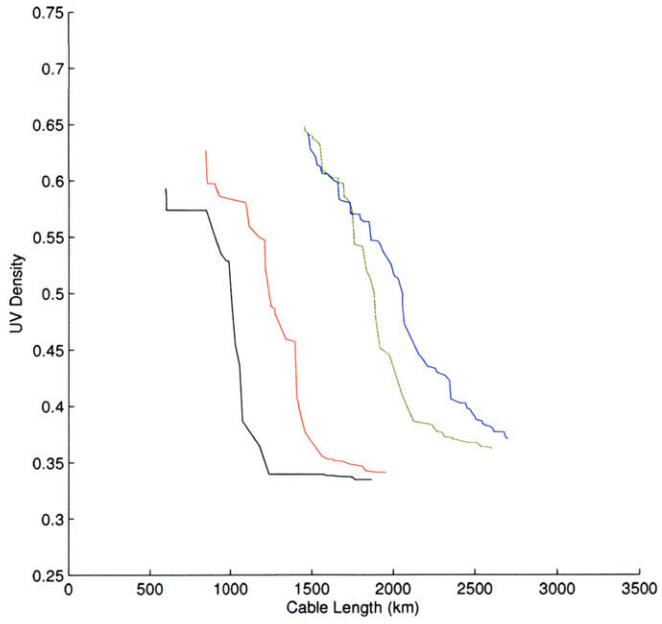


Figure 5-21: 160-station objective space varying *percent*, 4km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

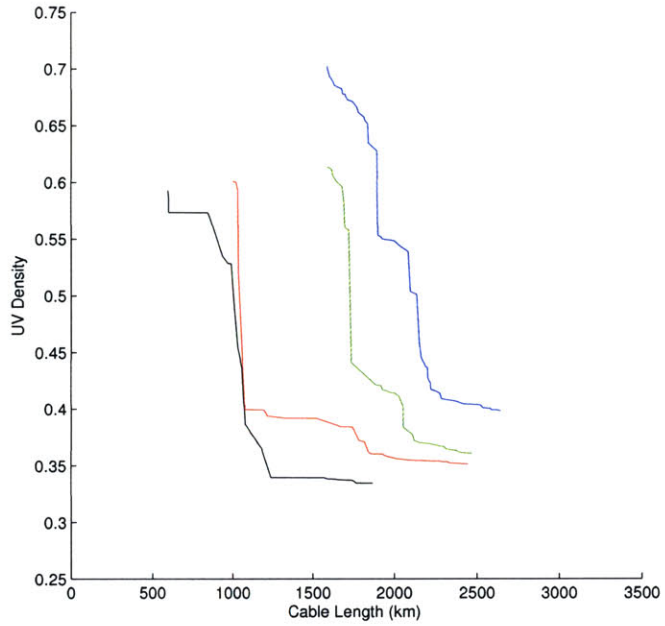


Figure 5-22: 160-station objective space varying *percent*, 20km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

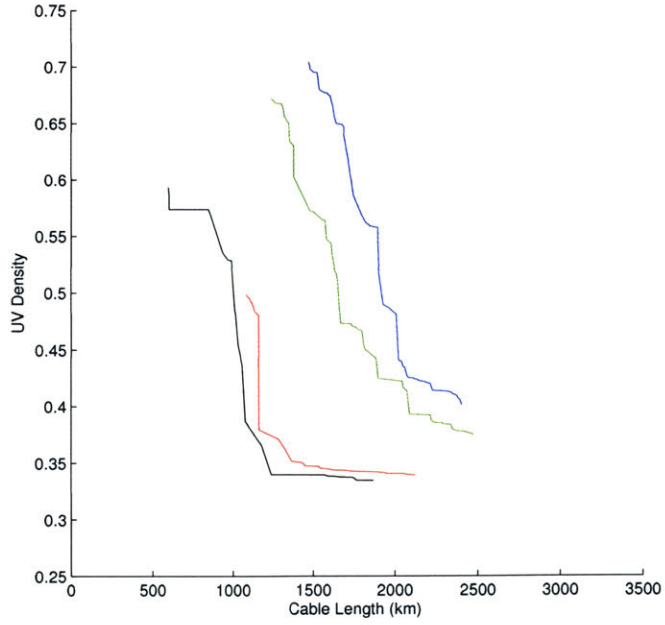


Figure 5-23: 160-station objective space varying *percent*, 40km *gridstep*. Unconstrained case is shown in black, 10% badzones in red, 30% badzones in green, and 50% badzones in blue.

Effect of *Gridstep* Parameter

Figures 5-24 - 5-35 show plots of the Pareto fronts while varying the site parameter *gridstep*. Plots were made for different *percent* sizes and number of stations. As can be seen from the plots the *gridstep* parameter does not affect the objective space as much as the *percent* parameter did. A noticeable effect is that as the number of stations increased, the Pareto fronts for the constrained cases got further and further away from the Pareto fronts of the unconstrained case. This shows that increasing *gridstep* makes it harder to utilize the entire xy plane. When *percent* was changing, all parts of the xy plane were still accessible, but as *gridstep* increases in size, entire swaths of xy plane are removed from the available variable space. This shows a coupling of varying *gridstep* with *percent*. As *percent* and *gridstep* increase, the swaths combine to mask out larger and larger areas, greatly limiting the useable xy plane. This makes it so that the relative performance metric is degraded because many of the baselines become more clumped in the uv plane. This also increases the cable length needed because connectivity of isolated station groups will require more of the longer cable pathways.

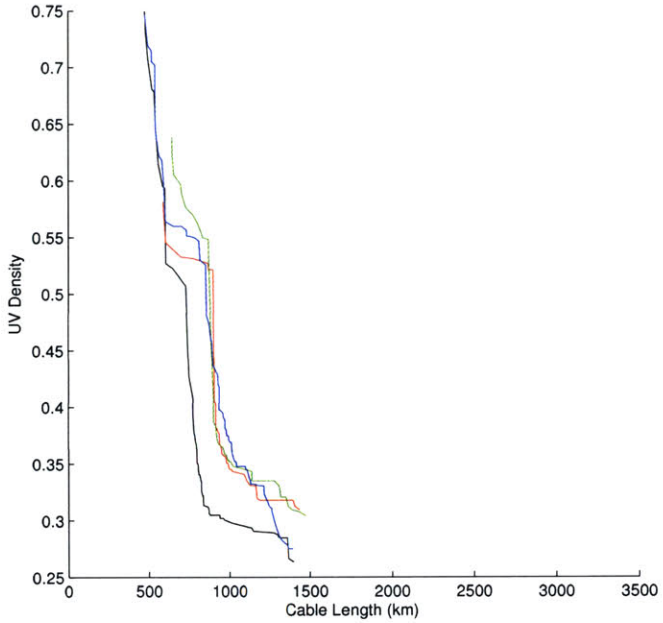


Figure 5-24: 27-station objective space varying *gridstep*, 10% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

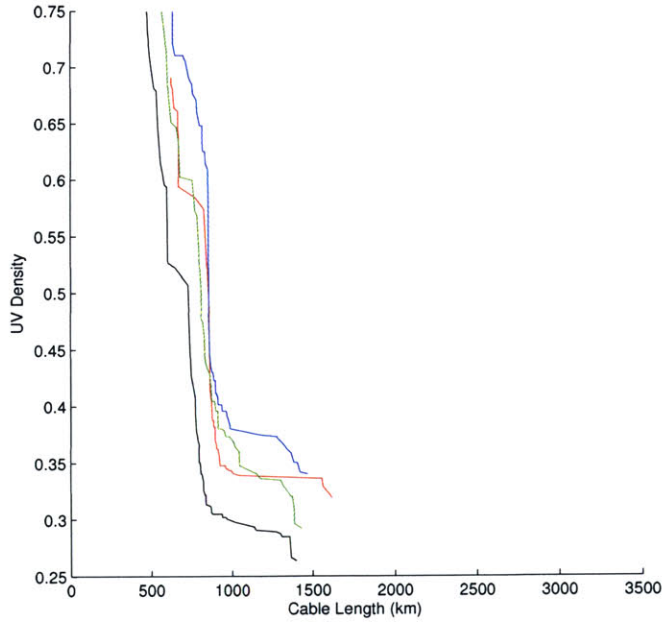


Figure 5-25: 27-station objective space varying *gridstep*, 30% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

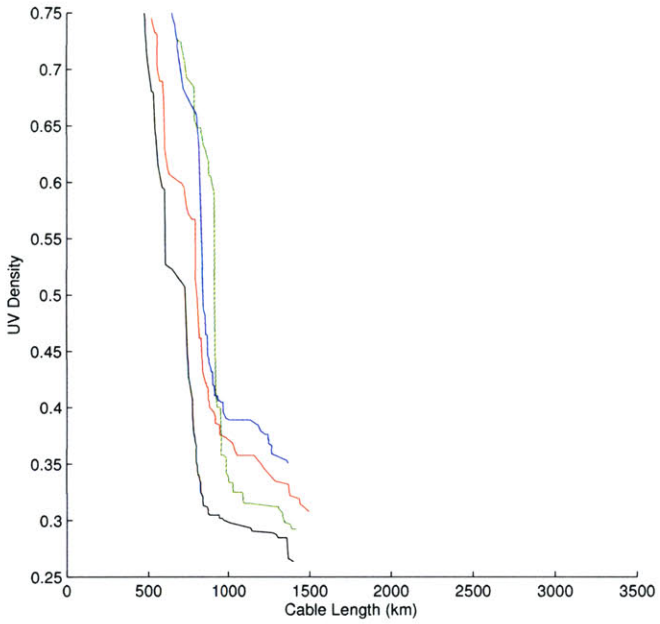


Figure 5-26: 27-station objective space varying *gridstep*, 50% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

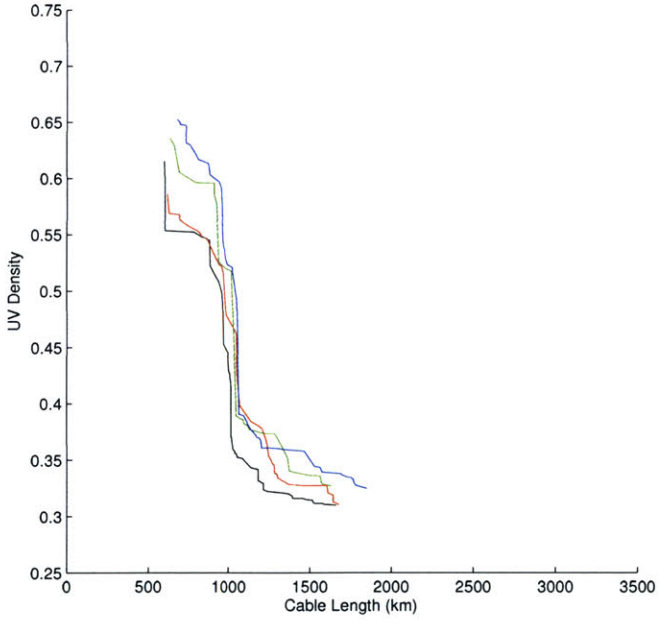


Figure 5-27: 60-station objective space varying *gridstep*, 10% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

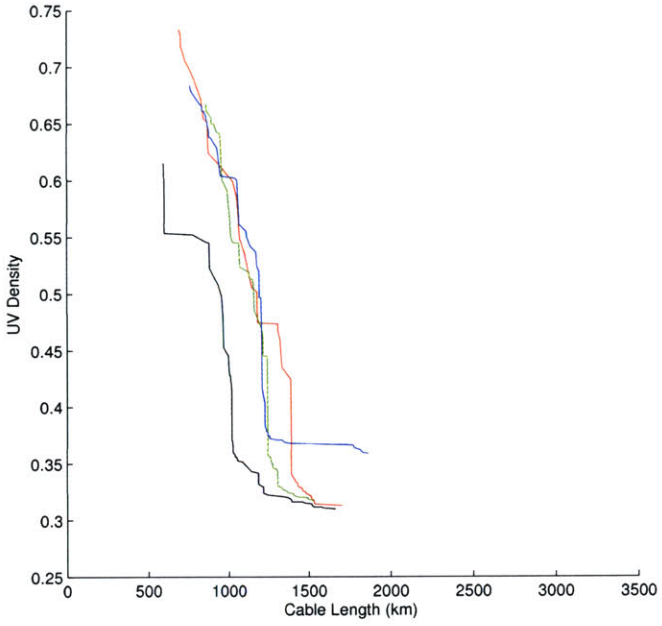


Figure 5-28: 60-station objective space varying *gridstep*, 30% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

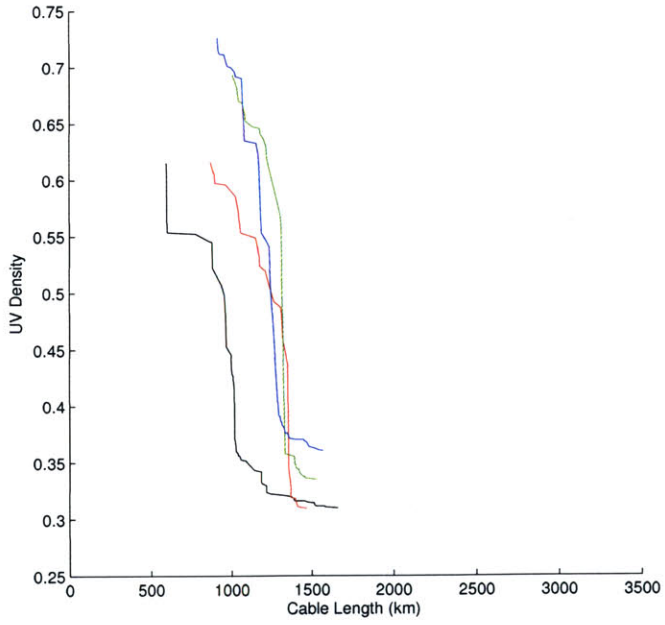


Figure 5-29: 60-station objective space varying *gridstep*, 50% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

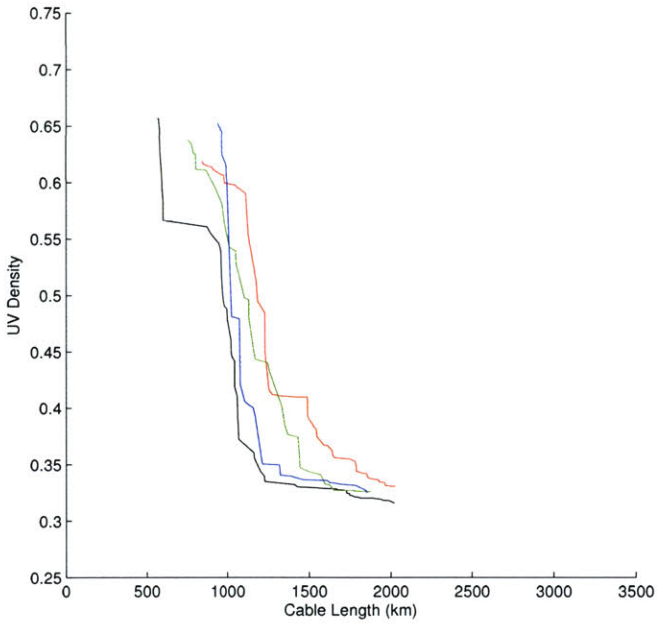


Figure 5-30: 100-station objective space varying *gridstep*, 10% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

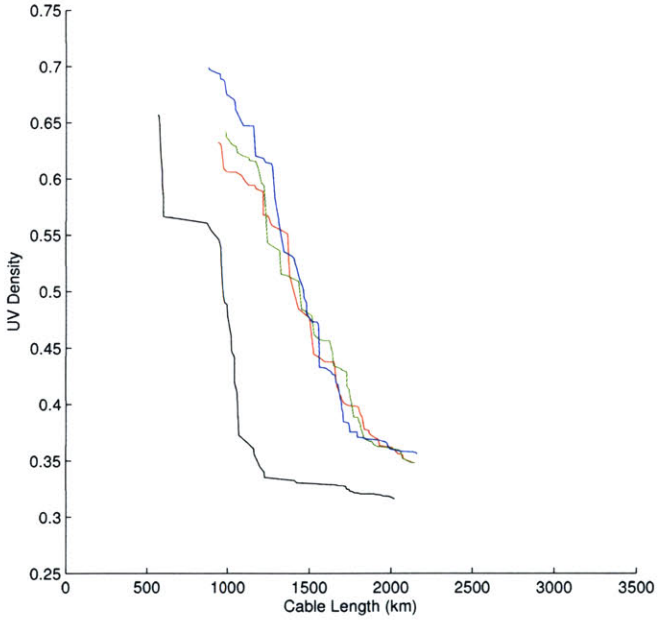


Figure 5-31: 100-station objective space varying *gridstep*, 30% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

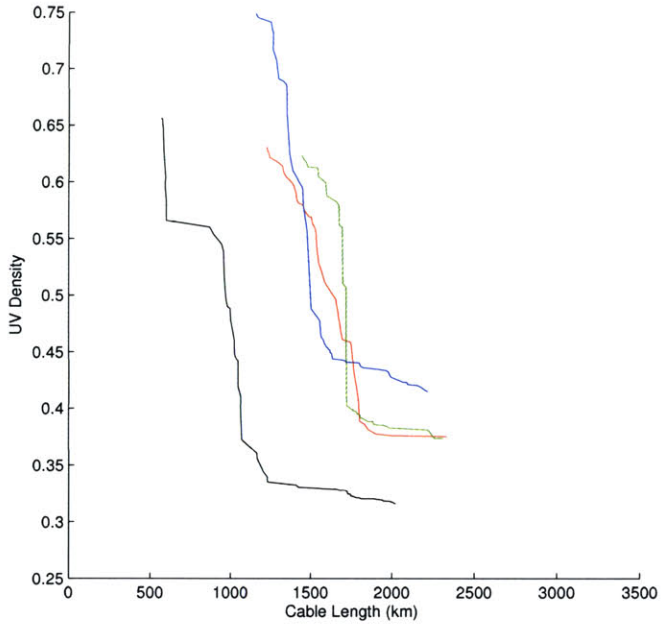


Figure 5-32: 100-station objective space varying *gridstep*, 50% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

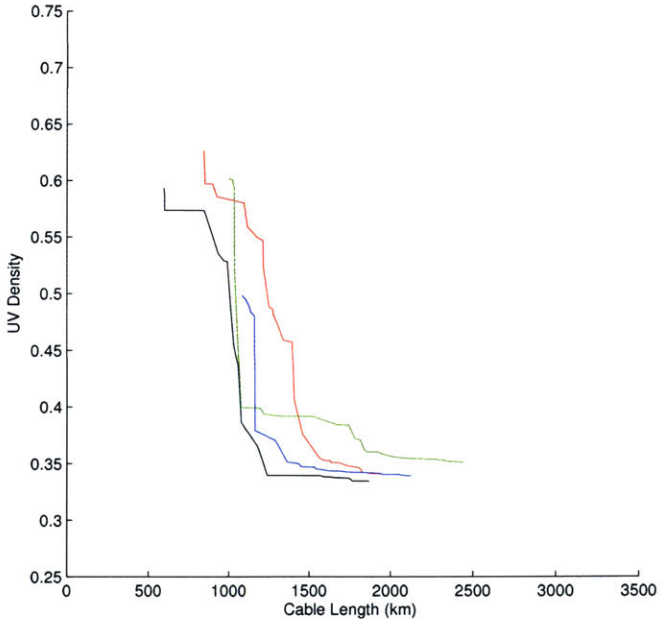


Figure 5-33: 160-station objective space varying *gridstep*, 10% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

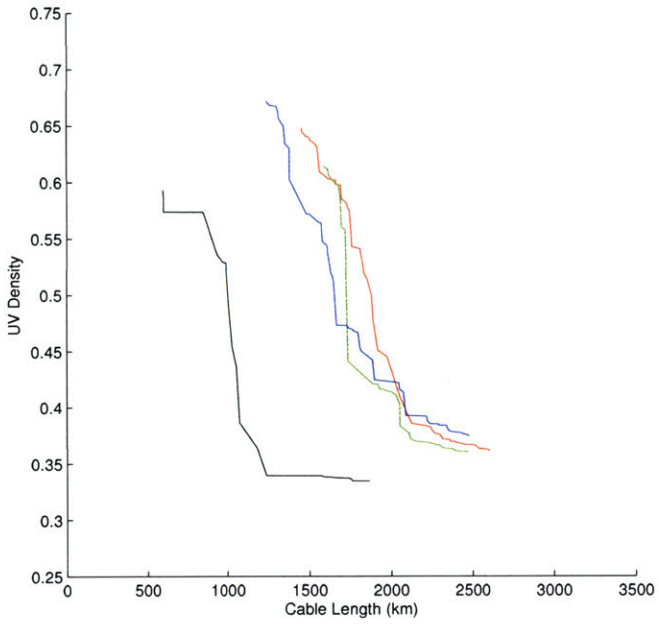


Figure 5-34: 160-station objective space varying *gridstep*, 30% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

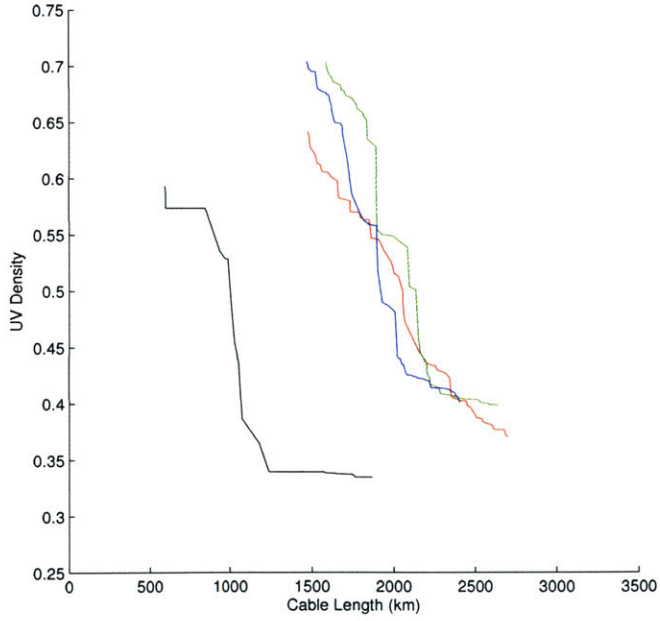


Figure 5-35: 160-station objective space varying *gridstep*, 50% badzones. Unconstrained case is shown in black, 4km *gridstep* in red, 20km *gridstep* in green, and 40km *gridstep* in blue.

5.3 Geometric vs. Non-Geometric Initial Seeds

Perturbations from ideal geometries and reduction in unnecessary components of the highly geometric initial seeds were sufficient to improve the designs. Section 5.2.1 discusses the possibilities of why no new topologies were found. To compare highly geometric initial seeds to non-geometric initial seeds simulations were run with only non-geometric initial seeds. Figure 5-36 shows the objective space for an unconstrained 60-station simulation that was run with only non-geometric initial seeds. As can be seen from the figure, the non-geometric initial seeds start significantly off the Pareto front. Figure 5-37 shows the difference between the Pareto fronts for the unconstrained 60-station simulation run with geometric initial seeds vs the unconstrained 60-station simulation run with non-geometric initial seeds. As can be seen, both converge to the high performance solutions similarly. As one moves along the Pareto front to the highly geometric designs (up and to the left) the non-geometric initial seed simulation did not populate that area in the Pareto front. Figure 5-38 shows the anchor and nadir-utopia solutions from the non-geometric simulation. As can be seen, the solutions do not vary a great deal. This strengthens the entropy argument that it is difficult to produce highly geometric seeds from non-geometric ones. This is why the initial seeds into the population are so important.

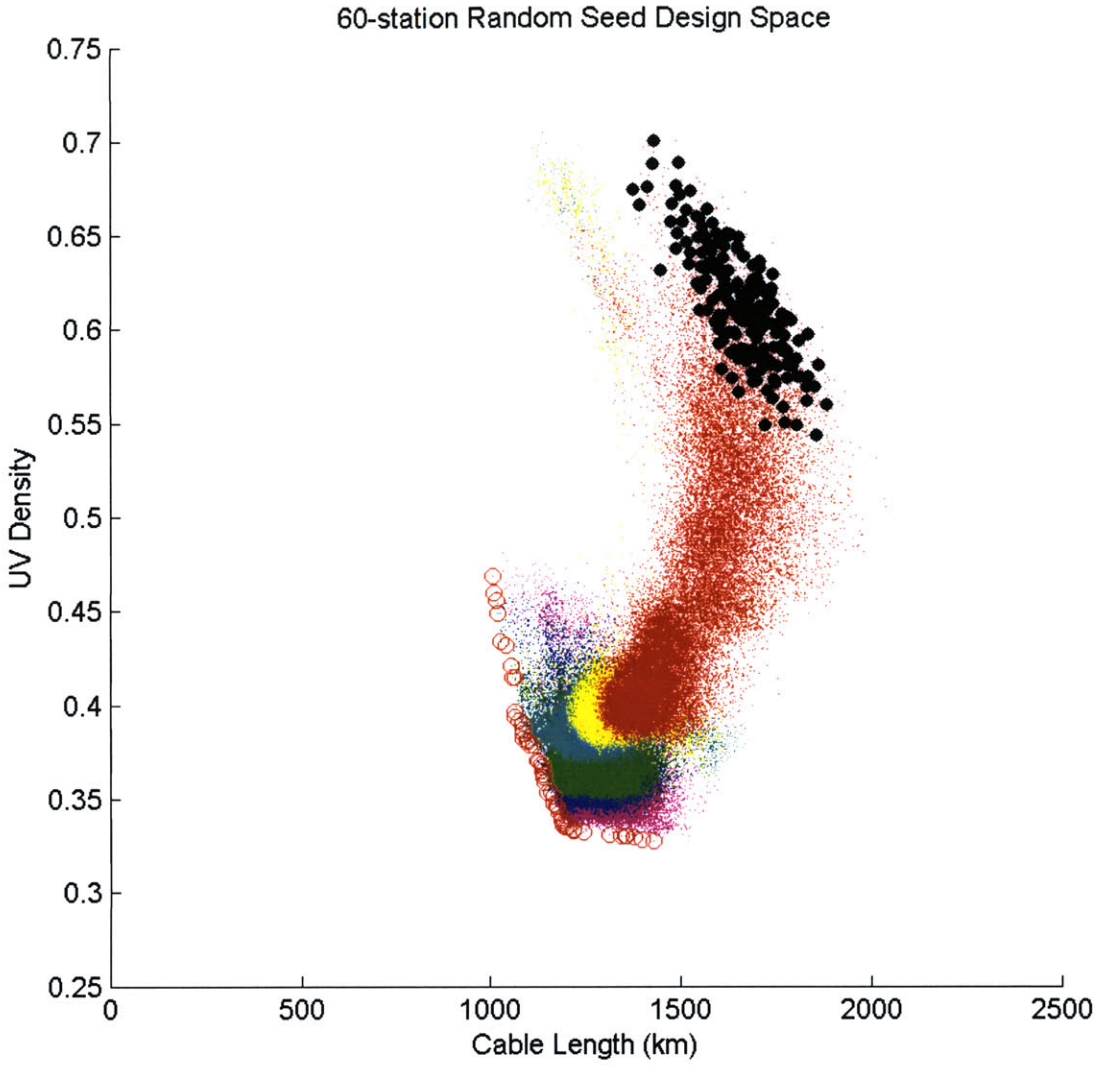


Figure 5-36: Unconstrained 60-station objective space with non-geometric initial seeds. Black dots denote initial seeds of the population. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

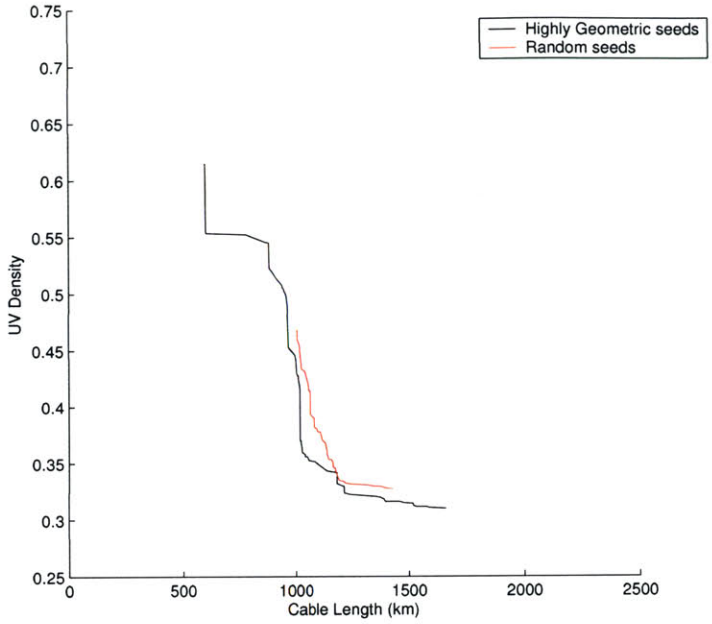


Figure 5-37: Comparison of non-geometric vs. geometric initial seeds Pareto fronts for unconstrained 60-station cases. The black line corresponds to the geometric while the red line corresponds to the non-geometric case.

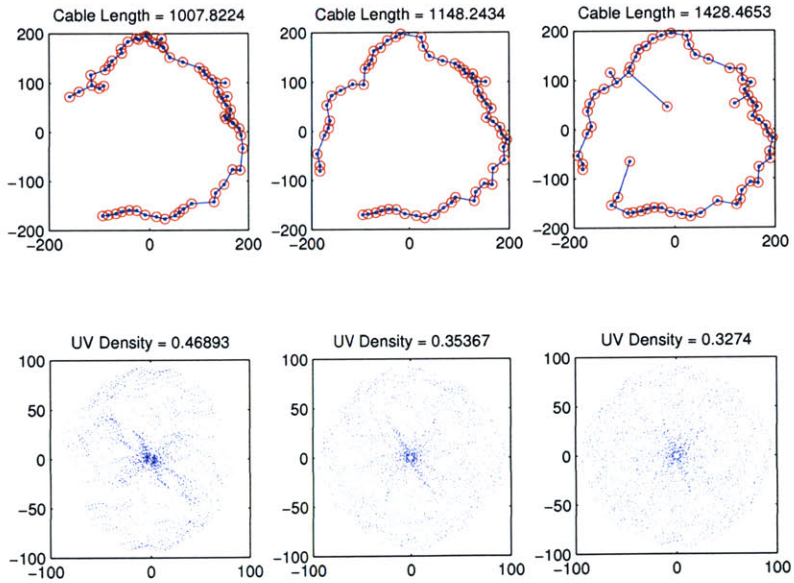


Figure 5-38: Unconstrained 27-station configurations (top) with corresponding uv coverage (bottom) for non-geometric initial seeds. Minimum cable configuration (left), nadir-utopia configuration (center), and maximum performance configuration (right).

5.4 Chapter 5 Summary

Chapter 5 discusses convergence and objective space issues in the multiobjective optimization framework. Issues are raised about convergence information pertaining to the multiobjective optimization framework. Objective spaces are explored and compared as a function of the number of stations, $N_{stations}$, as well as the site parameters of *gridstep* and *percent*. Non-geometric initial seed simulations are compared with geometric initial seed simulations. Results show that site constraints can greatly affect the overall design of radio telescope arrays and definitely need to be considered.

Chapter 6

Conclusion

This final chapter will summarize previous points and address future work, as well as show an example of a possible method of comparing configurations across number of stations. Simulated sky brightness maps using a number of different configurations will be created and compared to the VLA [20] to show just how much improvement can be expected from the next generation of radio interferometric telescopes. A summary of conclusions from designing the framework and running the case study based on LOFAR [14] design objectives, as well as future topics will be discussed briefly.

6.1 Beam Shapes and Imaging

In Chapter 1 the formulation for creating the point spread function of a beam from the xy station placements was briefly discussed. A comparison of beam shapes, like that of sidelobes in Figure 1-2, can be used to compare beams of any configuration, regardless the number of stations used. Briefly given here in Figures 6-1 - 6-6 are several beam comparisons as an example of how configurations can be compared. It is possible that configurations with a smaller number of stations can outperform configurations with a larger number of stations. This comparison was not done as a main part of the study for computational reasons. Computing the PSF from the xy station placements is too computationally expensive to do in this framework with the present hardware available.

As can be seen from Figures 6-1 - 6-6, the point spread functions, beam widths, and maximum sidelobes vary greatly depending on what configuration is chosen. This is just an example and was not exhaustively pursued, but gives a taste of what could be done with greater computational resources and the pursuit of other types of metrics.

A quantitative measure is to find the ratio of the maximum of the beam divided by the root-mean-square of the beam sidelobes. The details of the computation are not included here but a Matlab routine that calculates xy, uv, and PSF parameters has been attached in Appendix B. Table 6.1 gives these ratios for the six configurations given in Figures 6-1 - 6-6. As can be seen from the results, a configuration with greater stations does not necessarily perform better in this metric. A measure like this can be used to compare configurations across the parameter $N_{stations}$ in post-processing.

Configuration	Max PSF: RMS Sidelobes
60-station VLA	44.3
60-station nadir-utopia	78.2
60-station best performance	129.1
100-station VLA	45.0
100-station nadir-utopia	84.6
100-station best performance	147.4

Table 6.1: The ratio of maximum PSF value to RMS sidelobes for given configurations given in Figures 6-1 - 6-6 as a comparison study for other possible performance metrics.

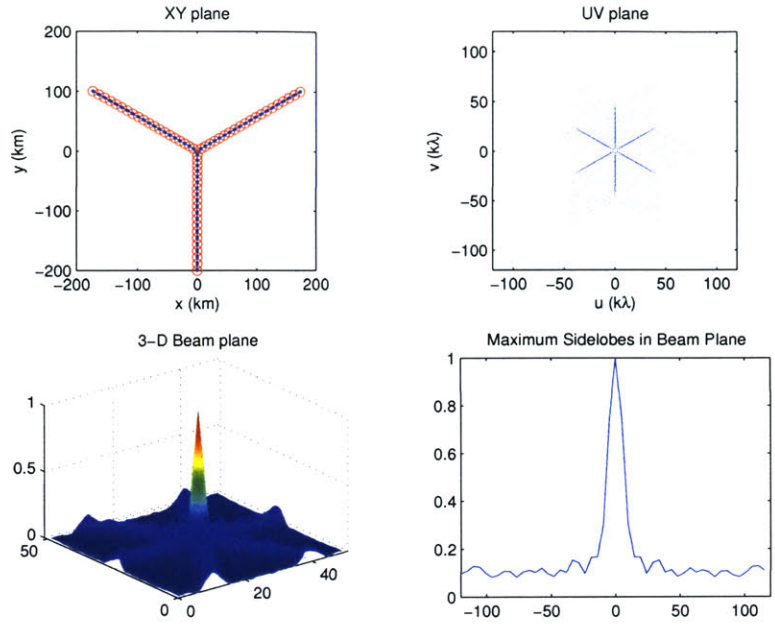


Figure 6-1: XY, UV, PSF, Sidelobe plot of 60-station VLA configuration. 60-station VLA configuration (upper-left) with its corresponding uv distribution (upper-right) and point spread function beam shape (lower-left) and maximum normalized beam sidelobes (lower-right).

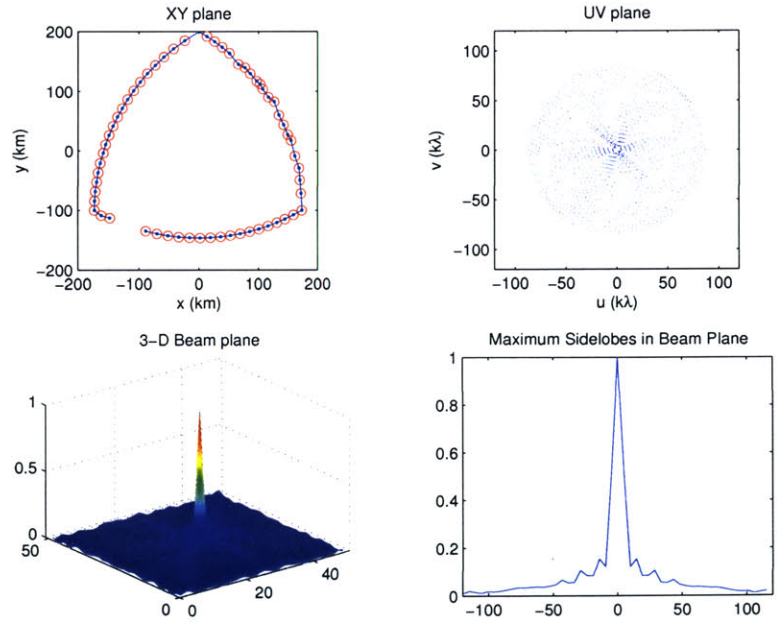


Figure 6-2: XY, UV, PSF, Sidelobe plot of 60-station nadir-utopia configuration. 60-station nadir-utopia configuration (upper-left) with its corresponding uv distribution (upper-right) and point spread function beam shape (lower-left) and maximum normalized beam sidelobes (lower-right).

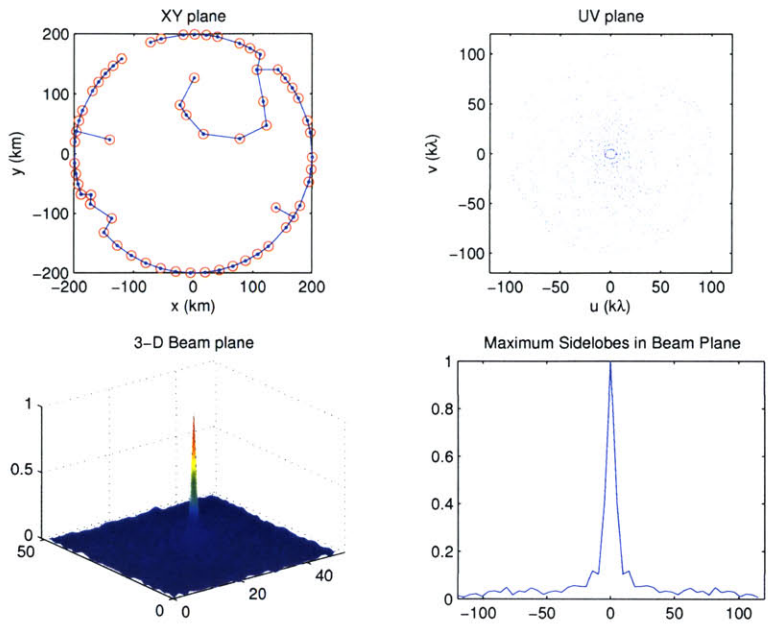


Figure 6-3: XY, UV, PSF, Sidelobe plot of 60-station best performance configuration. 60-station best performance configuration (upper-left) with its corresponding uv distribution (upper-right) and point spread function beam shape (lower-left) and maximum normalized beam sidelobes (lower-right).

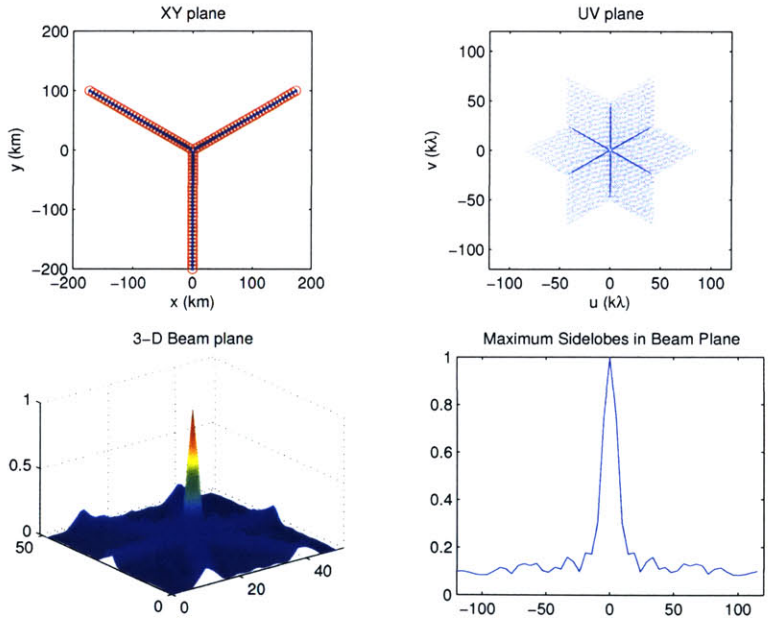


Figure 6-4: XY, UV, PSF, Sidelobe plot of 100-station VLA configuration. 100-station VLA configuration (upper-left) with its corresponding uv distribution (upper-right) and point spread function beam shape (lower-left) and maximum normalized beam sidelobes (lower-right).

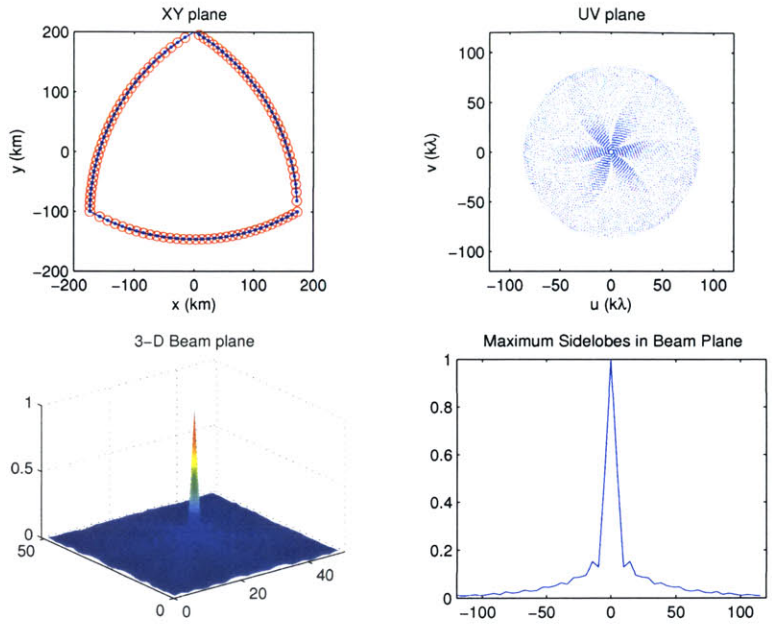


Figure 6-5: XY, UV, PSF, Sidelobe plot of 100-station nadir-utopia configuration. 100-station nadir-utopia configuration (upper-left) with its corresponding uv distribution (upper-right) and point spread function beam shape (lower-left) and maximum normalized beam sidelobes (lower-right).

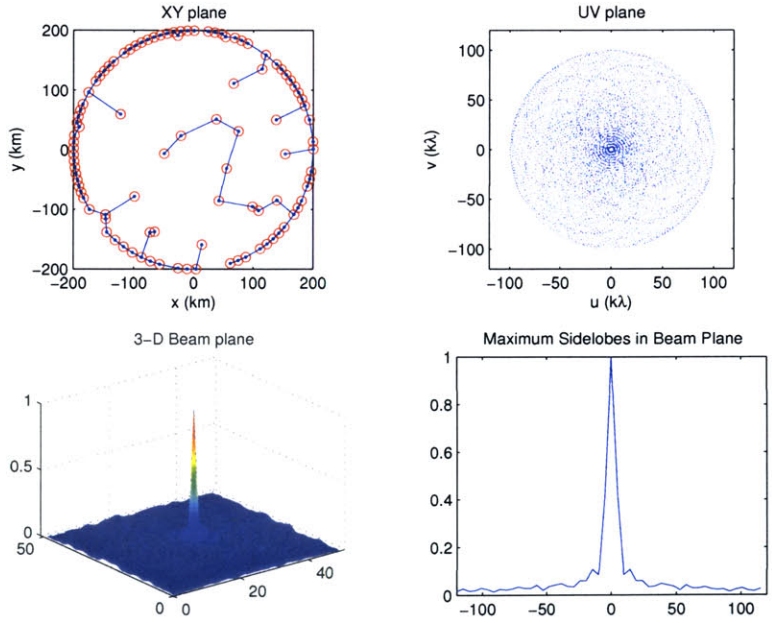


Figure 6-6: XY, UV, PSF, Sidelobe plot of 100-station best performance configuration. 100-station best performance configuration (upper-left) with its corresponding uv distribution (upper-right) and point spread function beam shape (lower-left) and maximum normalized beam sidelobes (lower-right).

6.2 Summary

In summary there are a number of things to say about the framework for optimization as well as results which have stemmed out of this study.

Initial Seeds are Important As in most optimizations, the initial seeds are very important for a number of reasons. First, they can drastically reduce the amount of time needed to converge to a final solution. Second, the use of heuristic algorithms can still leave parts of the objective space unexplored, especially when the problem is highly nonlinear, so initial seeds need to be chosen which will explore the desired parts of the objective space. Finally, initial seeds can be used to improve upon current configurations. In engineering there is usually a goal that is strived for in both performance and cost. Knowing and using previously good designs can aid in achieving these goals.

Site Constraints The need for site constraints is quite evident from all the results and comparisons between the unconstrained and constrained cases. Every real world system has constraints which it needs to deal with. These constraints can sometimes move optimal solutions far away from the ideal, unbound, theoretical case. For large ground-based arrays, site constraints are an important issue, but for any array based system, the constraints need to be addressed in a manageable fashion and having a framework that can handle this is very important. Historically, constraints have been difficult to deal with in genetic algorithms [10]. Hopefully this study has shown that it is manageable and computationally feasible to consider constraints.

Framework This framework is general enough to not only to address interferometric radio telescope arrays, but any large network that is affected by the placement of nodes for performance and cost. Adding additional metrics and more complex analysis tools will allow future users to expand upon research in this area. The example set here will hopefully bring the two worlds of science and engineering closer together so that both groups understand what needs to be done to make

the process of building these large complex systems simpler and more efficient.

More to Learn There is still a lot to learn about array configurations. In the next section on Future Work, topics will be discussed to expand upon the study of telescope arrays, networks, and the nature of multiobjective optimization. The field of complex system design now has the computational ability to tackle these types of harder problems and with this expanded ability comes new areas that can be explored which were not possible before.

6.3 Future Work

As this project matured it became evident that there were many avenues of research that could expand upon the usefulness of this framework. Presented here are many of those ideas.

$N_{stations}$ as a Design Parameter The number of stations can be a very significant cost in the design of radio telescope arrays. Minimizing this number while maintaining performance goals may help to drastically reduce cost. While this issue was not addressed in this case study formally, example results shown in this chapter suggest that this may be a real design parameter and will affect the optimization process greatly.

Complexity in Site Constraints The site constraints which were instituted here were simple in nature, allowing cabling to go anywhere, but limiting the placement of stations in the xy plane. In the future this can be expanded on in many ways. A few ideas include having different types of site constraints, ones which allow cabling, and ones which do not. Topography in both the site constraints would require changing the xy plane to the xyz plane, this increasing the number of design variables and complexity of the problem, but is possible and more realistic. This is also important for space-based arrays, which may require a three dimensional xyz plane.

Robustness Another interesting and important issue in the design of array systems is the robustness of the array. Any real system will have failures that occur in the elements of the array, degrading the overall performance of the array over time. Knowing which elements are more important than others, or which ones will cause catastrophic failures would aid in designing array configurations which have necessary redundancies or fail safes to litigate those types of problems.

Staged Deployment Many arrays are not built all at once and then just turned on, rather they are phased into existence. Research has been done on staged deployment [7] [19] and would be an interesting use of the framework to design arrays through time. This can be done in several ways. One way can be to know the end product that is desired, optimizing on that first, and then removing stations to work the problem backwards in time. The opposite way can be done as well. Given an initial number of stations, an optimum array can be built, and with each stage in the deployment, previous stages are frozen and the array is re-optimized with new stations being the only design variables.

Competing Science Objectives Arrays like the VLA [20] are reconfigurable because the scope of the telescope is broad and allows for research in multiple areas, thus adapting the properties of the array to different science objectives. A major issue between scientist and engineers is the adaptability of an array and how to accommodate the design process to build arrays with simultaneous competing science objectives.

Space-Based Arrays Space-based arrays can bring interesting case studies to the framework. Since they are in space there are different types of constraints and objectives to work with. The modelling also may need to be in a three dimensional plane if the array designers so desire. The framework is completely expandable to be able to handle these types of variables, constraints, and objectives, and would be an interesting application to explore.

Convergence for Multiobjective Optimization As described in Section 5.1.3,

the definition of convergence becomes ambiguous for multiobjective problems because the termination criteria may be poorly defined. A new method for finding termination criteria based upon objective space evolutionary information may be feasible. The objective space plots in this thesis show a progression of the array configurations from initial seeds to the final generation. It may even be possible to determine when a certain section of the objective space has converged and then refocus efforts and resources on different parts of the objective space which are still actively improving, such as is seen in the anchor solutions in this study.

New Optimization Methods Finally, highly nonlinear problems which have a large number of design variables and multiple objectives may be outmatching current optimization techniques. This can be seen through the difficulty of convergence, advancement past initial seeds, and destructive mating. A new method which is computationally rigorous and robust can possibly be developed to handle issues in sensitivity of design variables and objectives and adapt better to these highly nonlinear problems.

6.4 Final Remarks

I would just like to say that I am really glad for the opportunity to be able to work on this project. I would once again like to thank my advisors, Jackie and Oli, for all their help. There is a great advantage in having two advisors from two different fields. The perspectives and insights really helped in taking this thesis as far as it has gone.

I hope that this work can be developed into future projects described throughout the thesis, mainly in the section on Future Work, and that I will be involved in this development in some way or another.

Thank you once again to everyone.

The End

Bibliography

[1] Arecibo: http://www.naic.edu

[2] ATA: www.seti.org/science/ata.html

[3] Boone, F., 2001, Interferometric Array Design: Optimizing the Locations of the Antenna Pads, Astronomy & Astrophysics, 377, 368

[4] Boone, F., 2002, Interferometric Array Design: Distributions of Fourier Samples for Imaging, Astronomy & Astrophysics, 386, 1160

[5] Cornwell, T. J., Novel Principle for Optimization of the Instantaneous Fourier Plane Coverage of Correlation Arrays, 1988, IEEE Transactions on Antennas and Propagation, AP-36, 1165

[6] Darwin, C., 1859, On the Origin of Species, (Cambridge, MA: Harvard University Press)

[7] de Weck, O.L., de Neufville R., and Chaize M., Staged Deployment of Communications Satellite Constellations in Low Earth Orbit, Journal of Aerospace Computing, Information, and Communication, submitted November 8, 2003

[8] Dreyfus, S.E. & Wagner, R. A., The Steiner Problem in Graphs, 1971, Networks 1, 195, 207

[9] Golay, M., 1971, Point Arrays Having Compact Non-Redundant Autocorrelations, Journal of Optical Society America, Vol 61, pg 272

[10] Goldberg, D. E., 1989, Genetic Algorithms in Search, Optimization, & Machine Learning (Boston: Addison Wesley)

[11] Keto, E., Shapes of Cross-Correlation Interferometers, 1997, Astrophysical Journal, 475, 843

[12] Kogan, L., Optimization of an Array Configuration Minimizing Sidelones, 1997, ALMA Memo, 171

[13] Kogan, L., Optimization of an Array Configuration with a Topography Constraint, 1998a, ALMA Memo, 202

[14] LOFAR: www.lofar.org

[15] Pohlheim, H., 1997, GEATbx: Genetic and Evolutionary Algorithm Toolbox for use with Matlab, V1.92

[16] SKA: www.skatelescope.org

[17] Sneath, P. H. A., Application of Computers to Taxonomy, 1957, Journal of General Microbiology, 17, 201

[18] Strang, Curtis., 1986, Introduction to Applied Mathematics (Wellesley, MA: Wellesley Cambridge Press)

[19] Takeuchi, H. et al, 2000, 18th International Atomic Energy Agency Fusion Energy Conference, IAEA-CN-77-FTP2/03

[20] Thompson A. R., Clark B. G., Wade C. M., Napier P. J., The Very Large Array, 1980, Astrophysical Journal Supplements, 44, 151

[21] Thompson, A. R., Moran, J. M., Swenson, G. W., 1986, Interferometry and Synthesis in Radio Astronomy (New York: Wiley Interscience)

[22] Woody, D., 1999, ALMA Configurations with Complete UV Coverage, ALMA Memo, 270

[23] Woody, D., 2001, Classical Configuration Evaluation

[24] Zitzler, E., 2002, Methods for Design, Optimisation, and Control

Appendix A

Figures

A.1 Non-Uniform UV Distributions

Besides the uniform uv distribution there can also be skewed uv distributions to accommodate different types of science objectives. Figure A-1 shows nominal uv distributions for radial distributions of r^{-1} & r^{-2}.

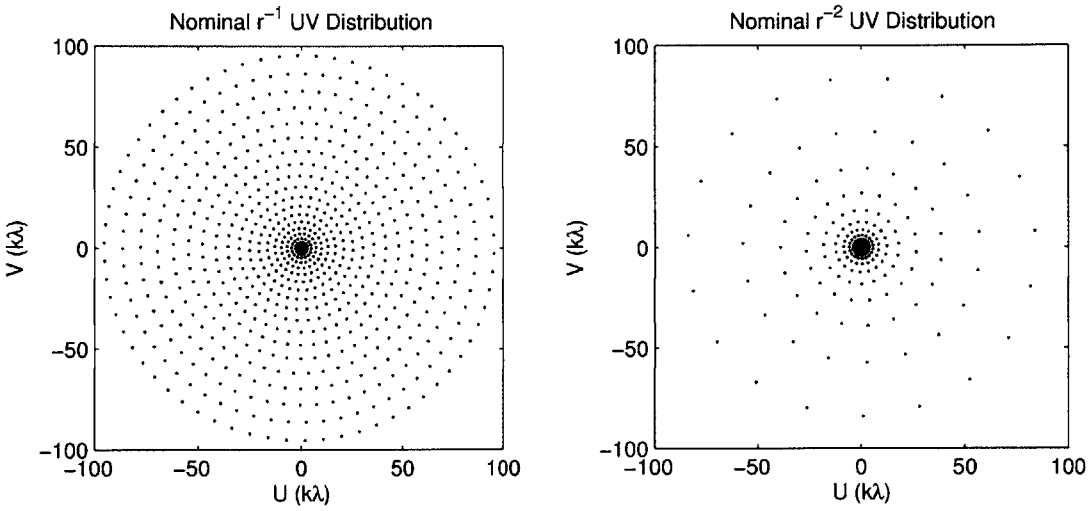
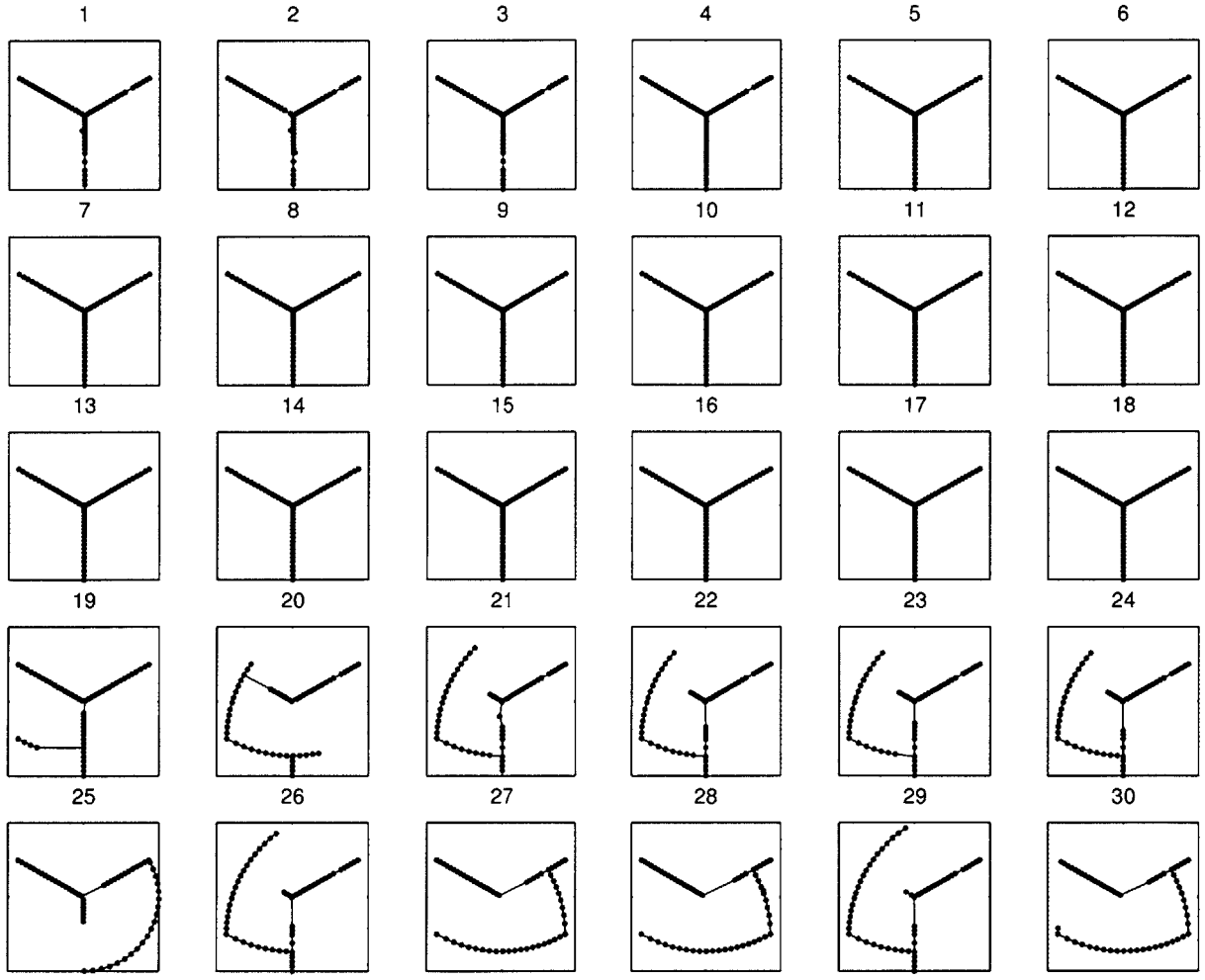
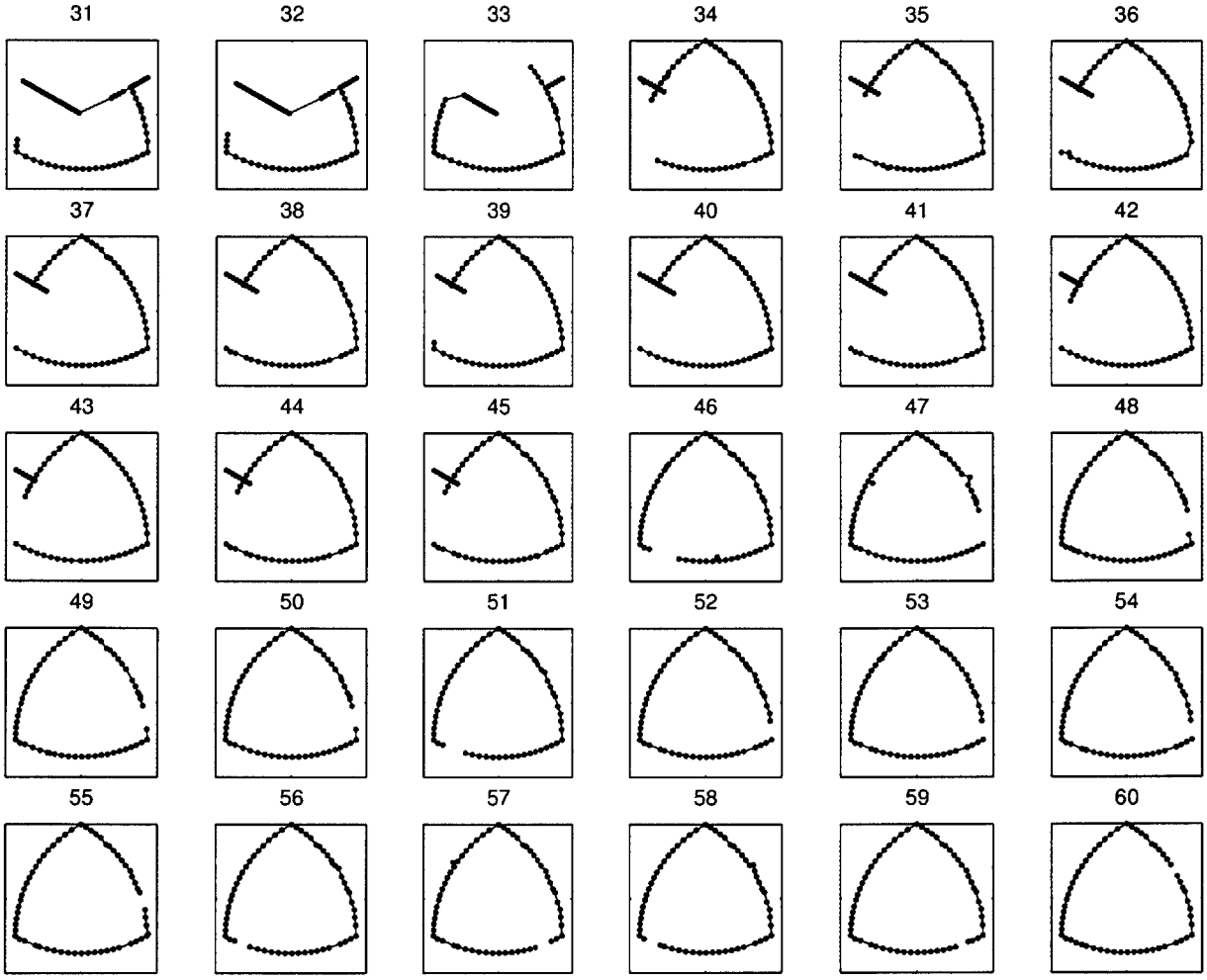


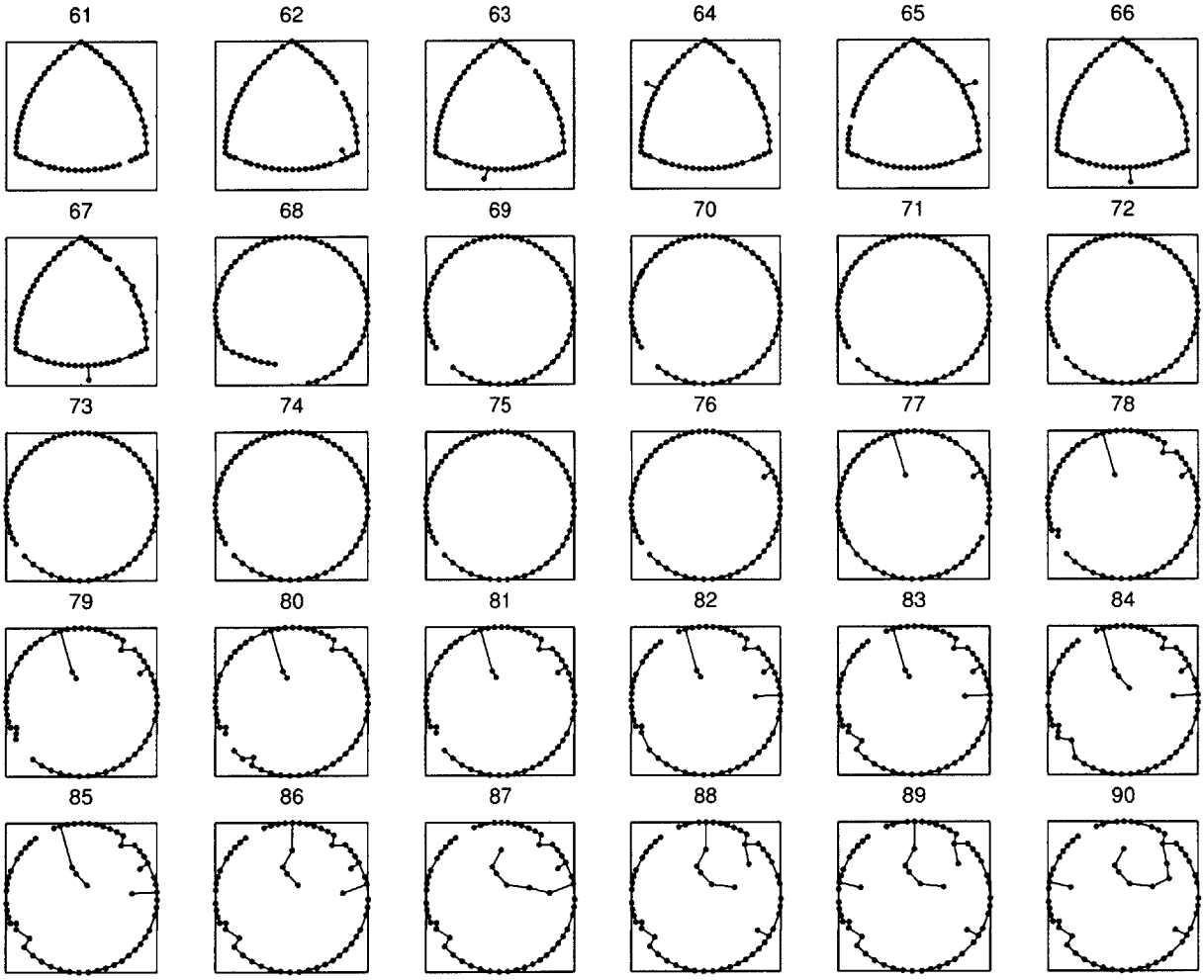
Figure A-1: Nominal r^{-1} & r^{-2} uv distribution for a 27-station ($2 * 351$ uv point) configuration.

A.2 60-station Pareto configurations

Here is shown a family of Pareto solutions corresponding to the non-dominated solutions in Figure 5-9. Configurations are sorted from the minimum cable anchor solution to the best performance anchor solution.




31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60


61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90

A.3 Convergence Plots

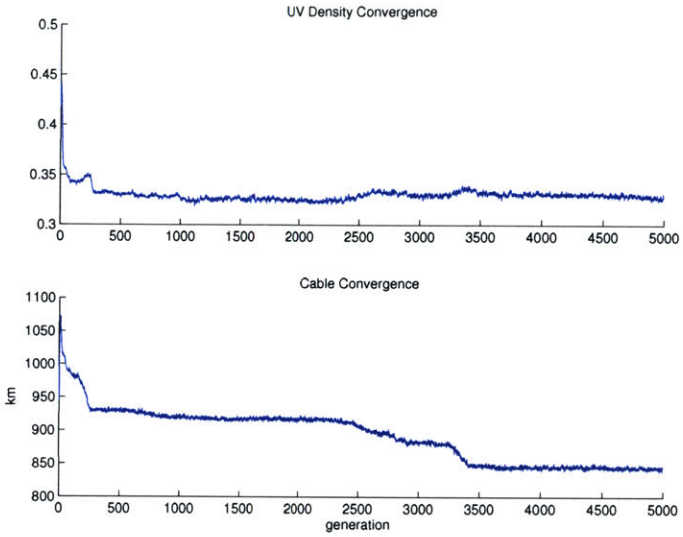


Figure A-2: Unconstrained 27-station convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

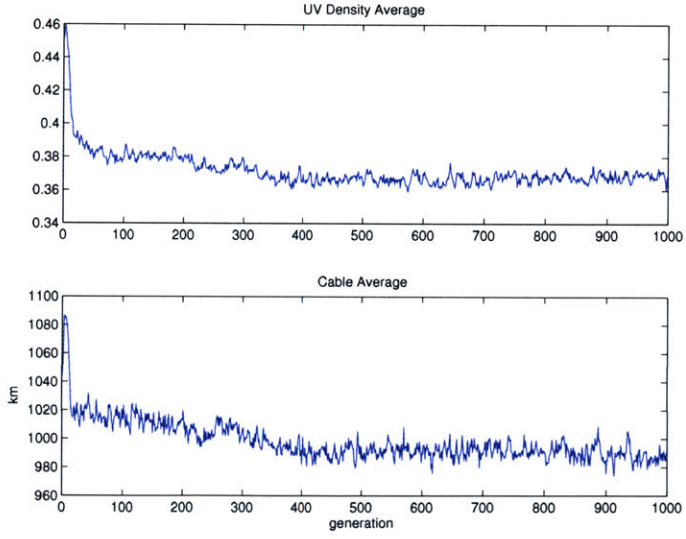


Figure A-3: 27-station, 4km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

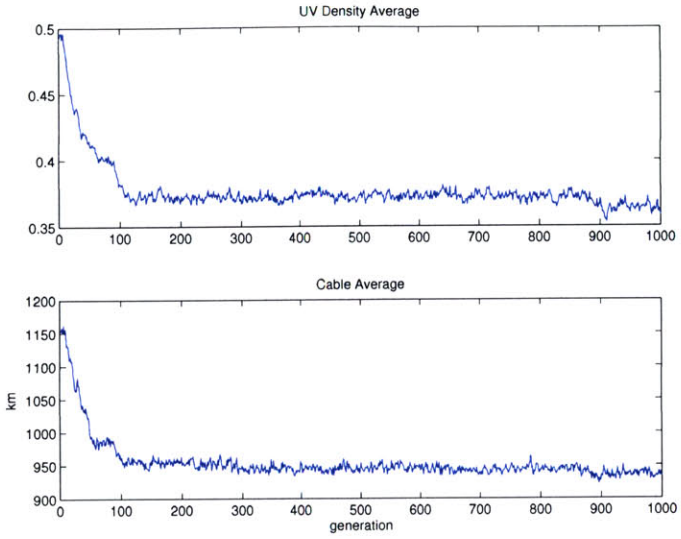


Figure A-4: 27-station, 4km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

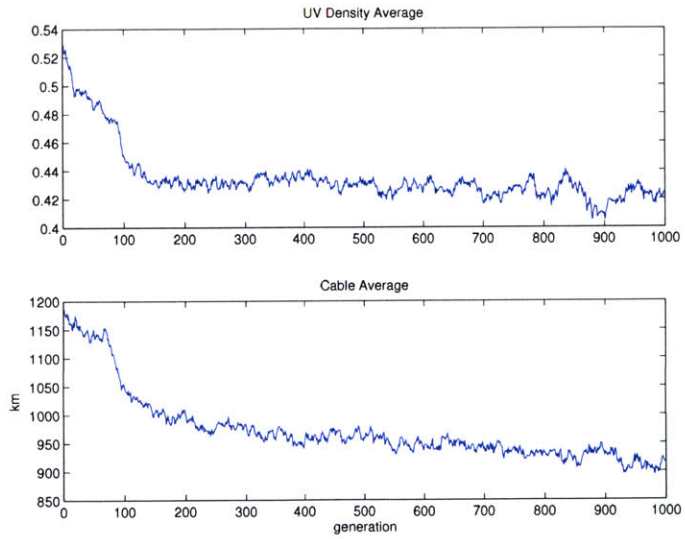


Figure A-5: 27-station, 4km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

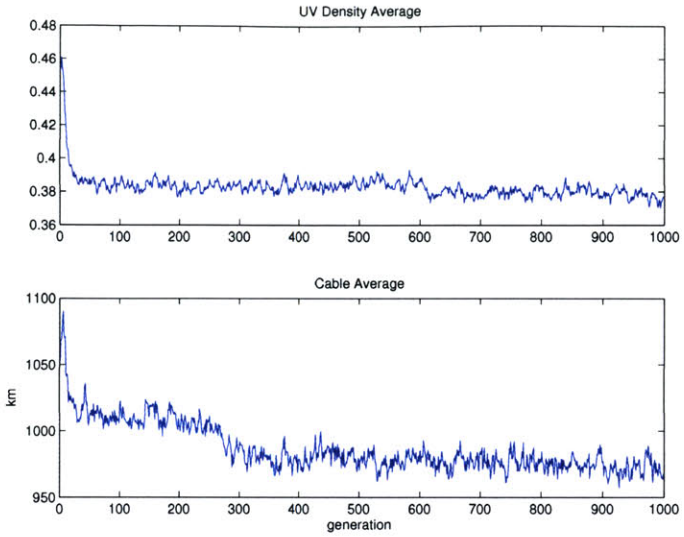


Figure A-6: 27-station, 20km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

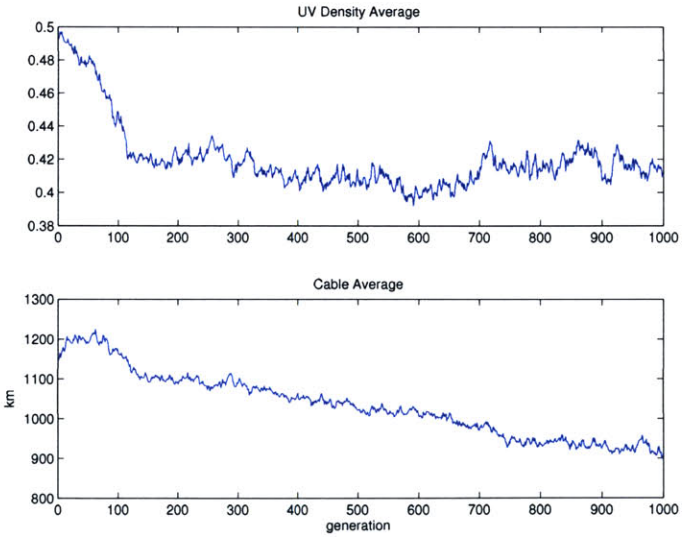


Figure A-7: 27-station, 20km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

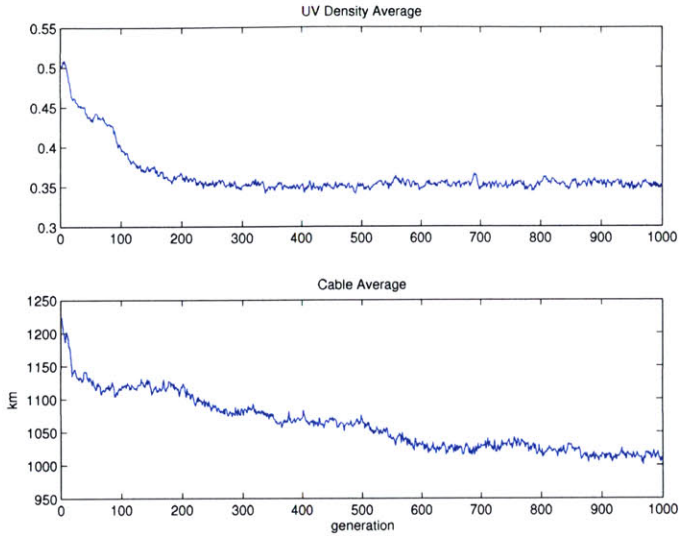


Figure A-8: 27-station, 20km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

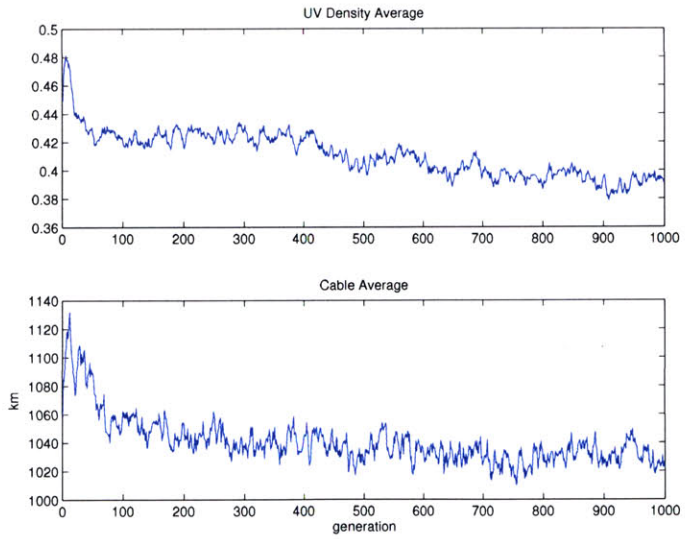


Figure A-9: 27-station, 40km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

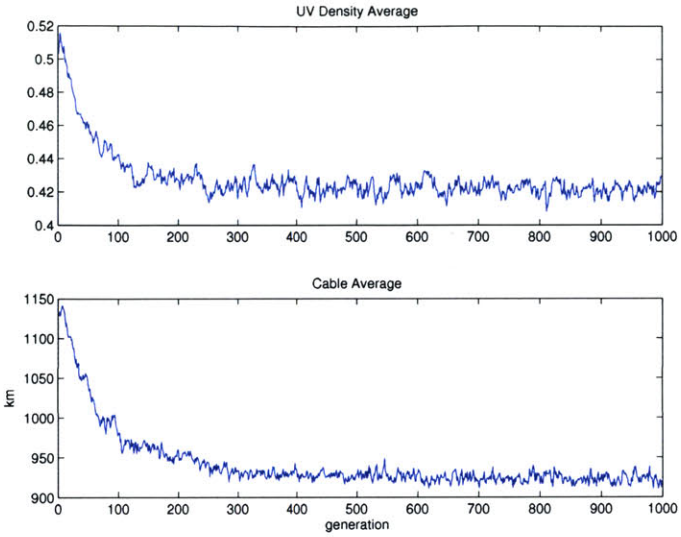


Figure A-10: 27-station, 40km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

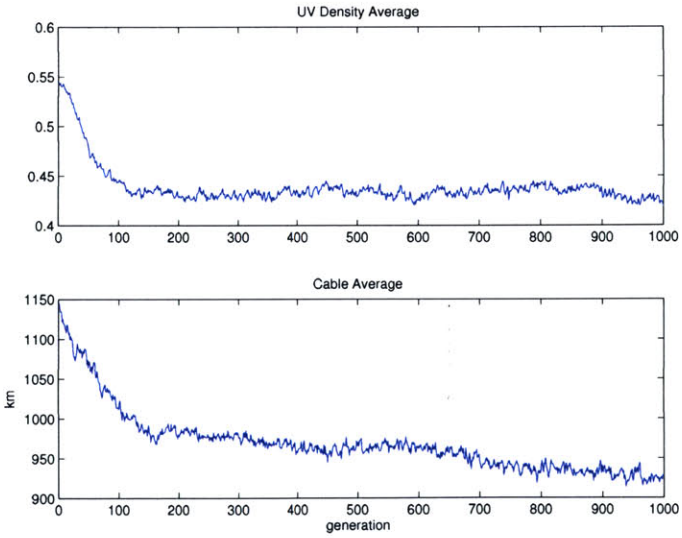


Figure A-11: 27-station, 40km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

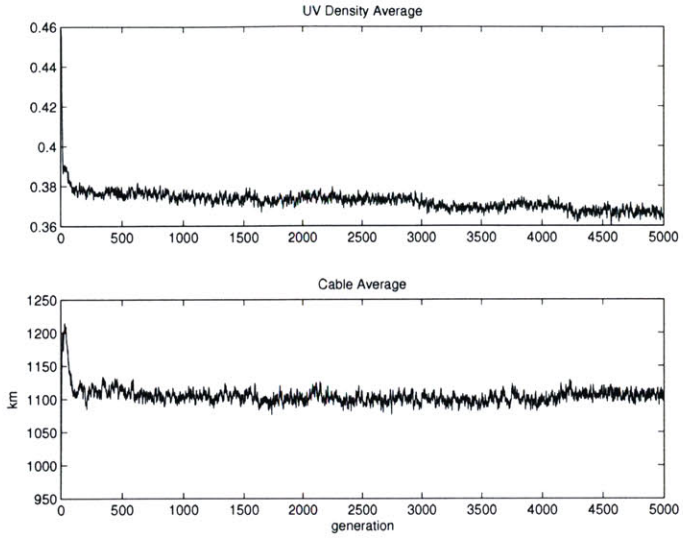


Figure A-12: Unconstrained 60-station convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

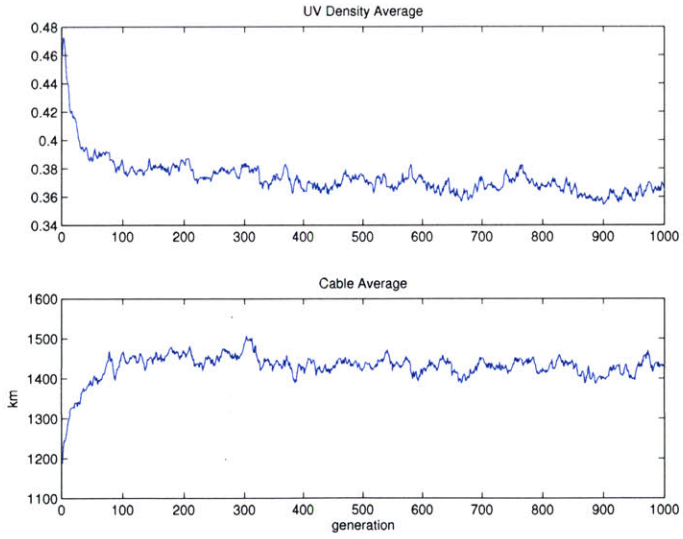


Figure A-13: 60-station, 4km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

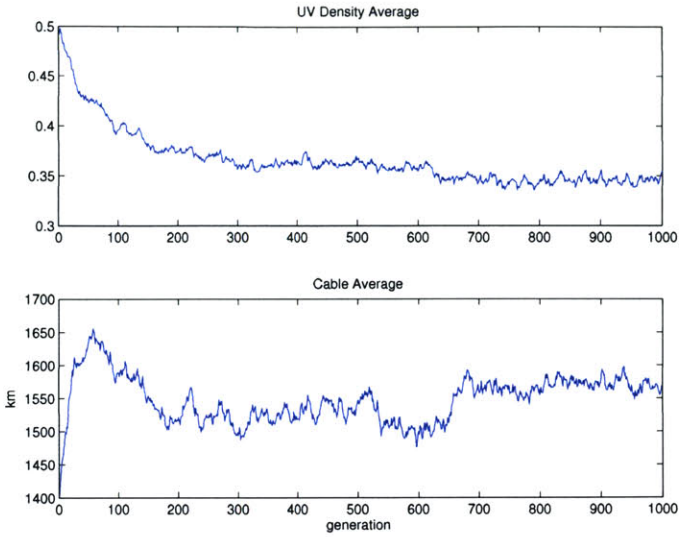


Figure A-14: 60-station, 4km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

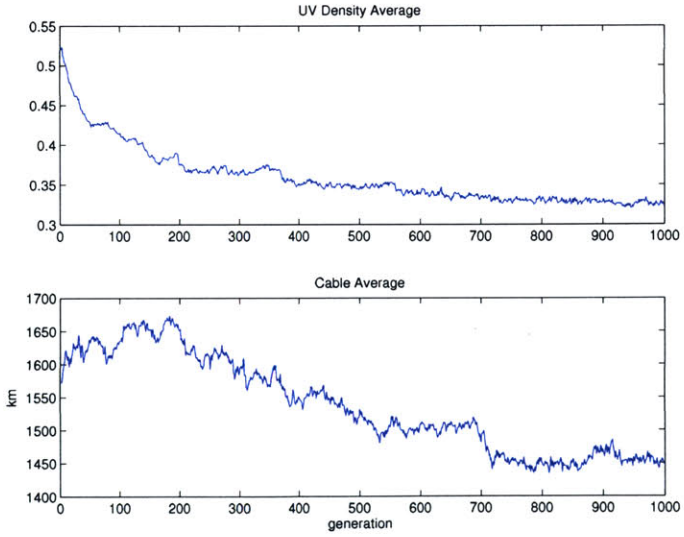


Figure A-15: 60-station, 4km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

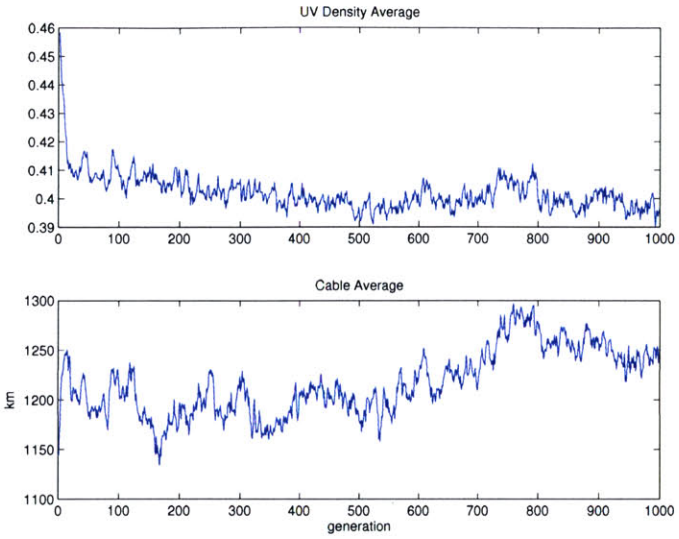


Figure A-16: 60-station, 20km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

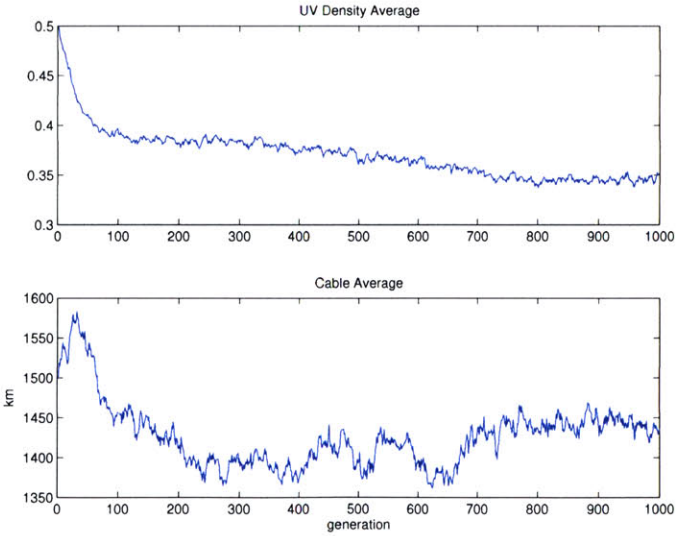


Figure A-17: 60-station, 20km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

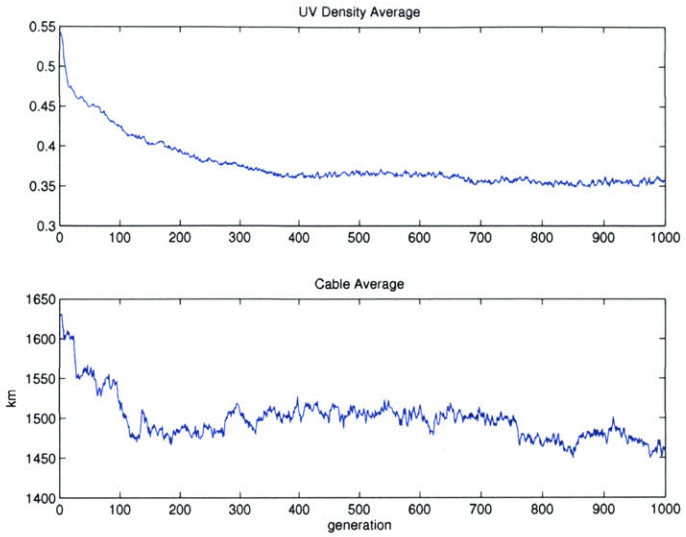


Figure A-18: 60-station, 20km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

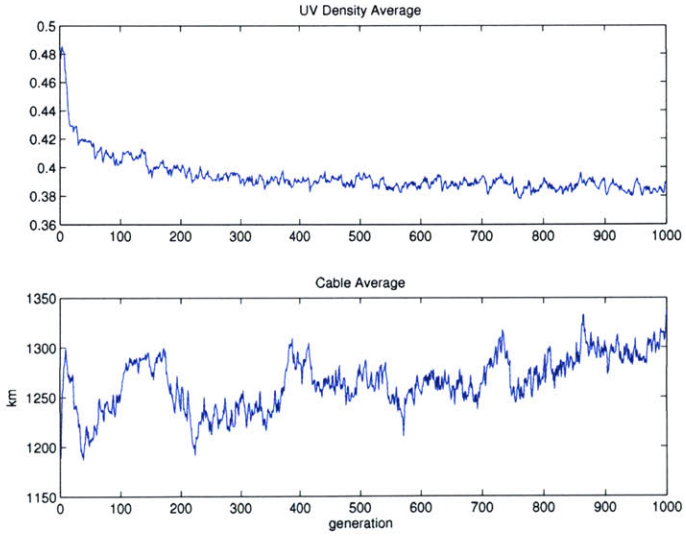


Figure A-19: 60-station, 40km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

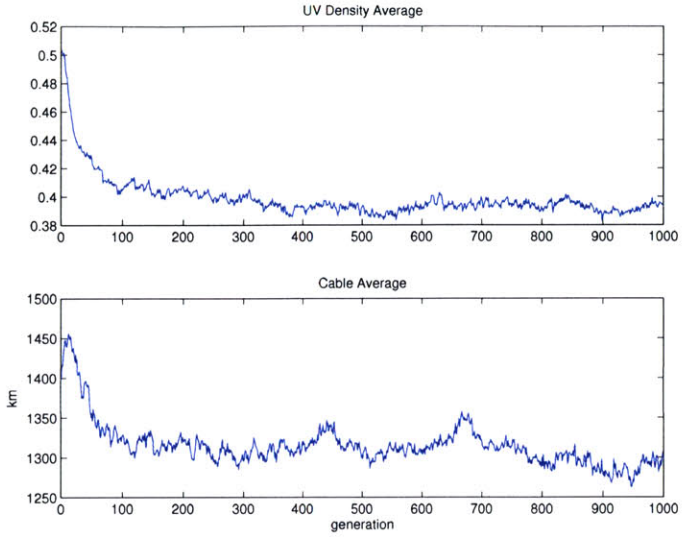


Figure A-20: 60-station, 40km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

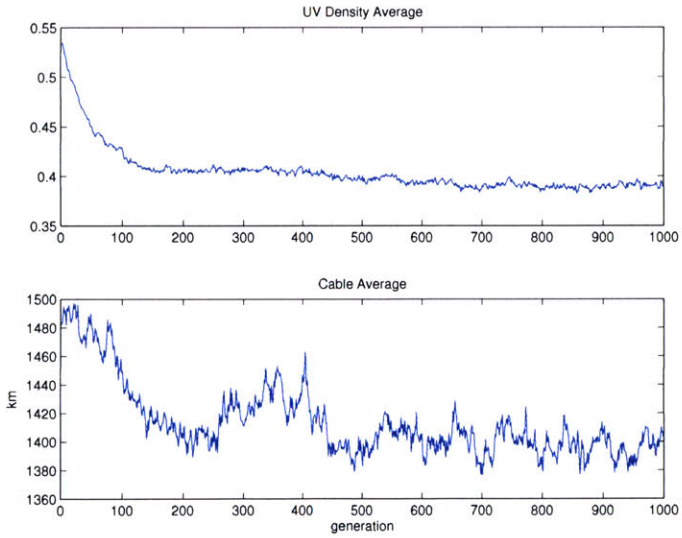


Figure A-21: 60-station, 40km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

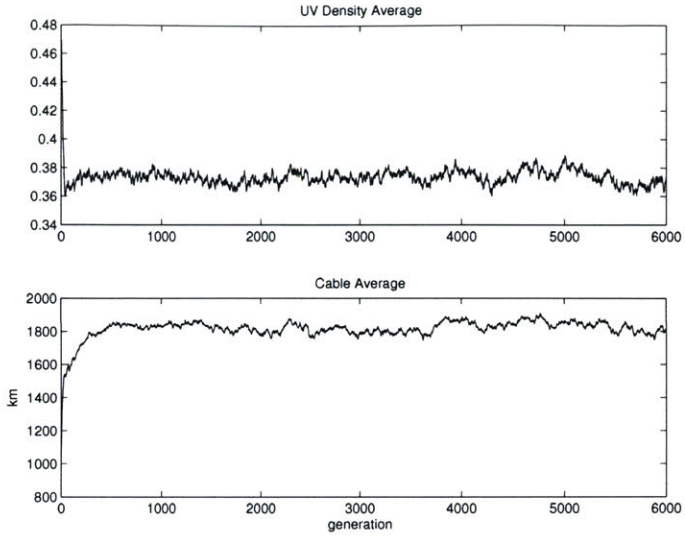


Figure A-22: Unconstrained 100-station convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

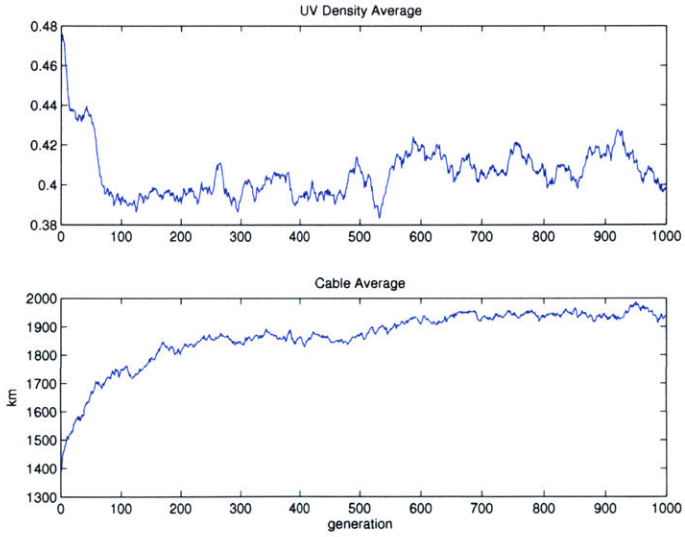


Figure A-23: 100-station, 4km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

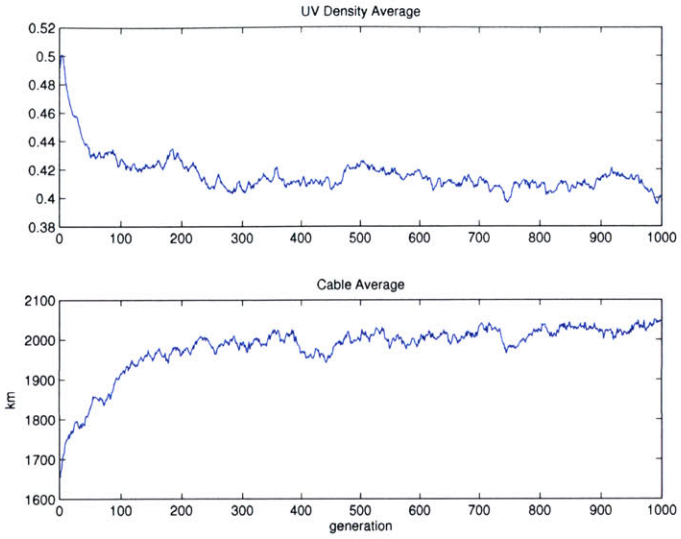


Figure A-24: 100-station, 4km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

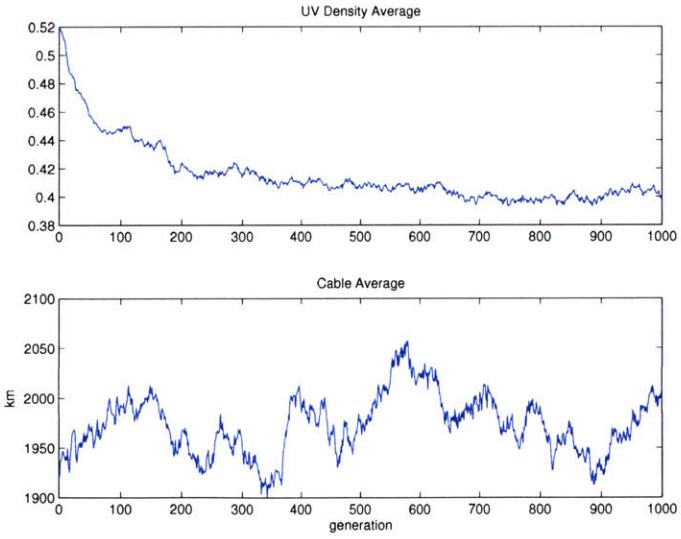


Figure A-25: 100-station, 4km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

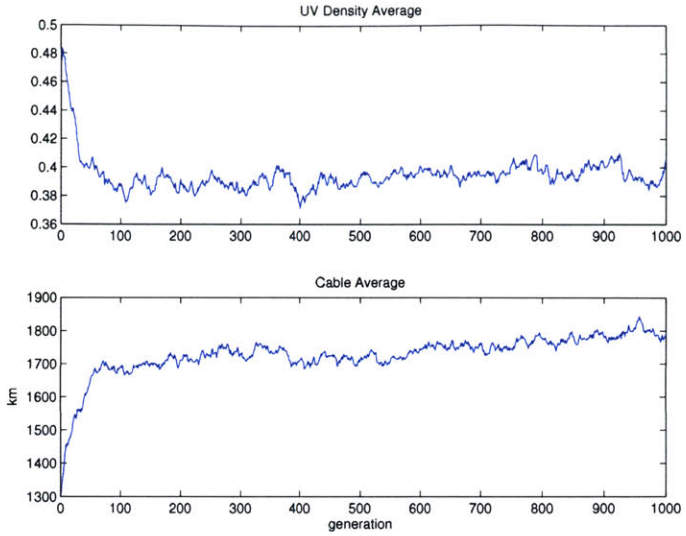


Figure A-26: 100-station, 20km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

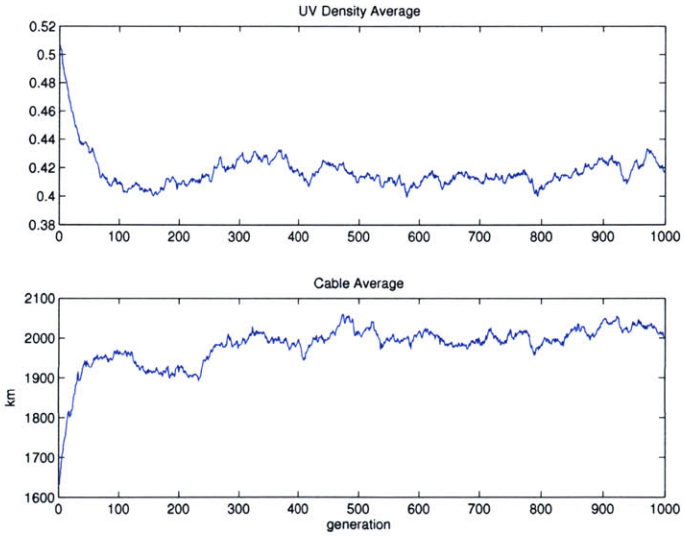


Figure A-27: 100-station, 20km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

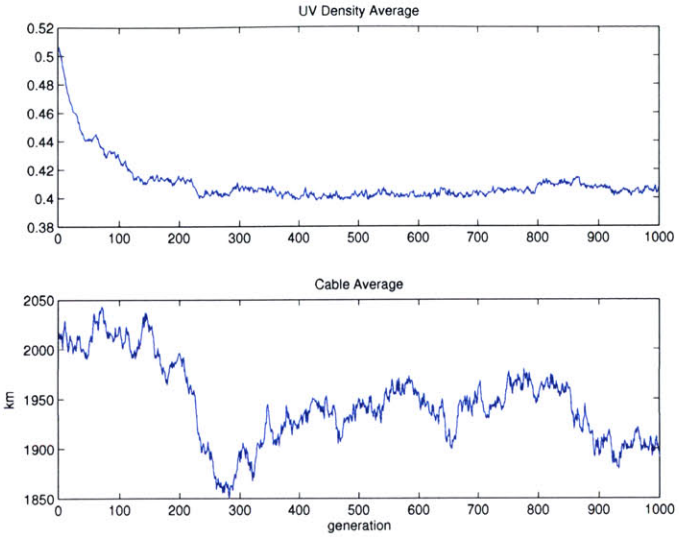


Figure A-28: 100-station, 20km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

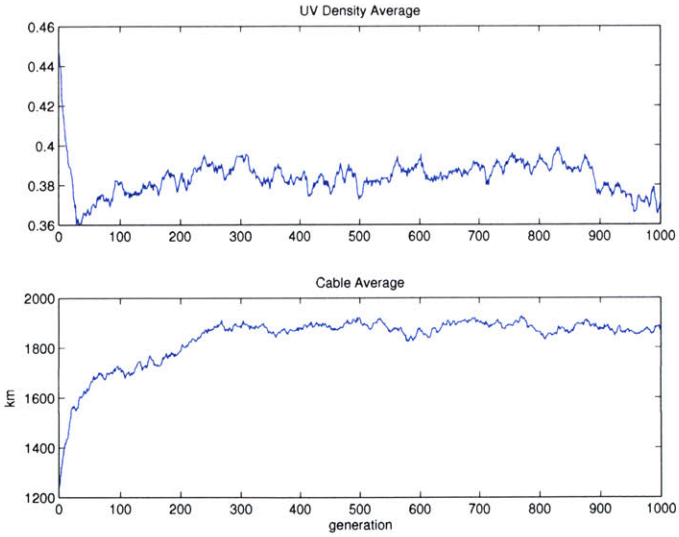


Figure A-29: 100-station, 40km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

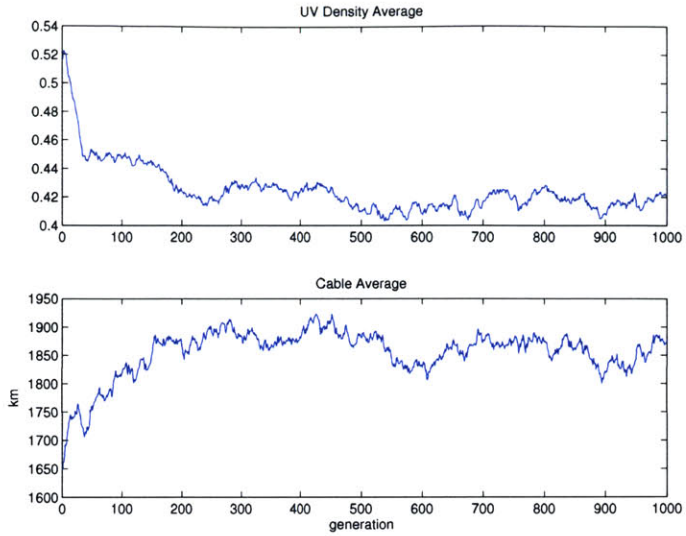


Figure A-30: 100-station, 40km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

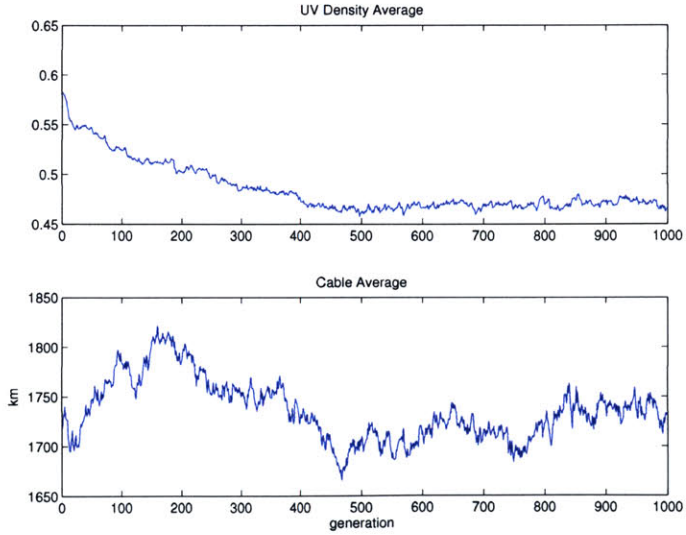


Figure A-31: 100-station, 40km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

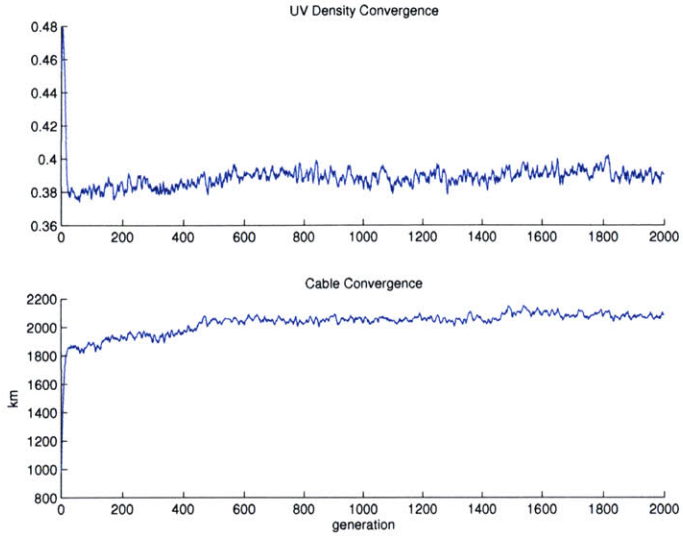


Figure A-32: Unconstrained 160-station convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

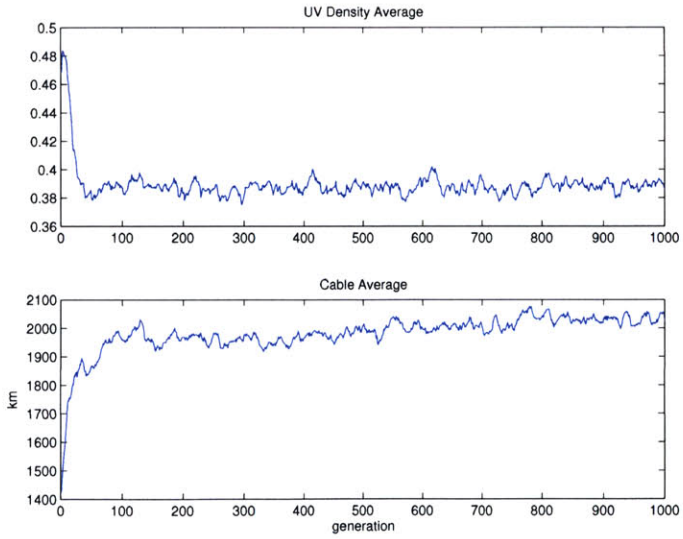


Figure A-33: 160-station, 4km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

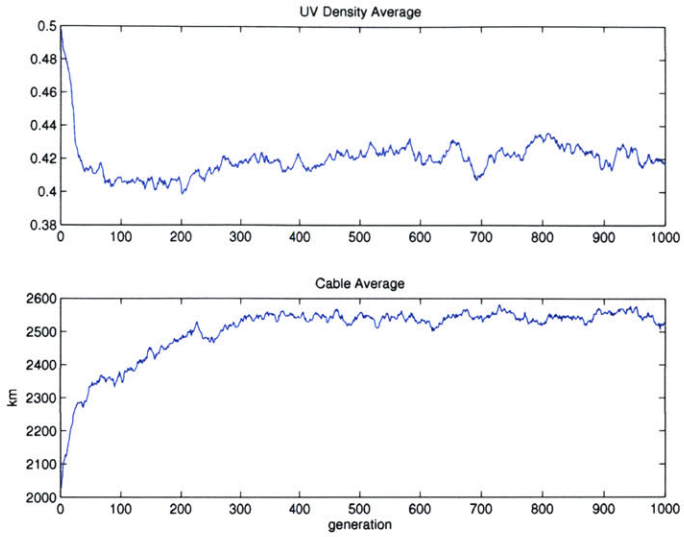


Figure A-34: 160-station, 4km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

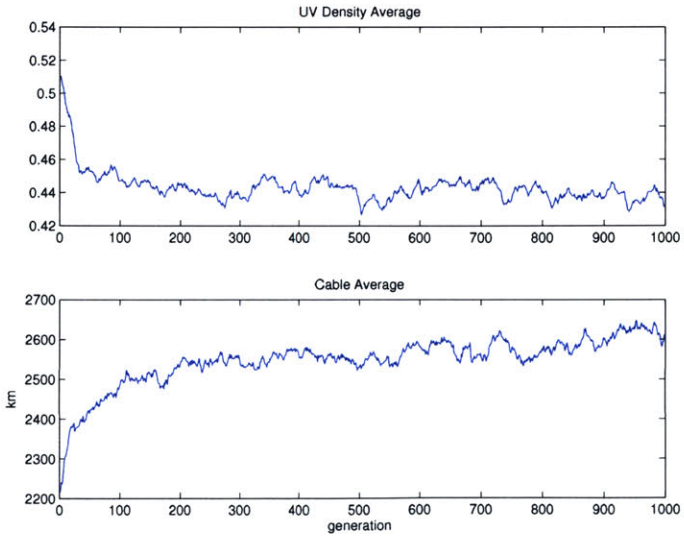


Figure A-35: 160-station, 4km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

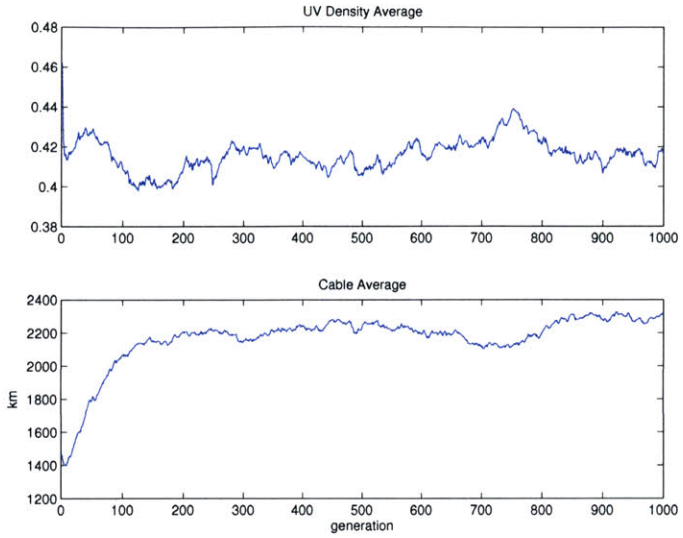


Figure A-36: 160-station, 20km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

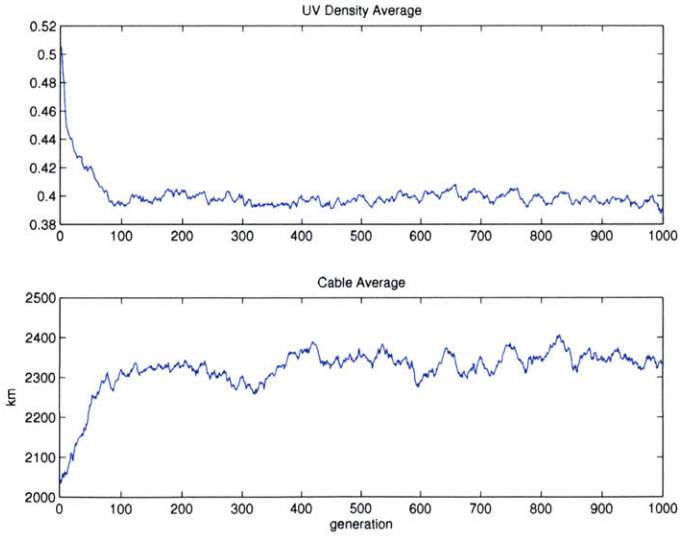


Figure A-37: 160-station, 20km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

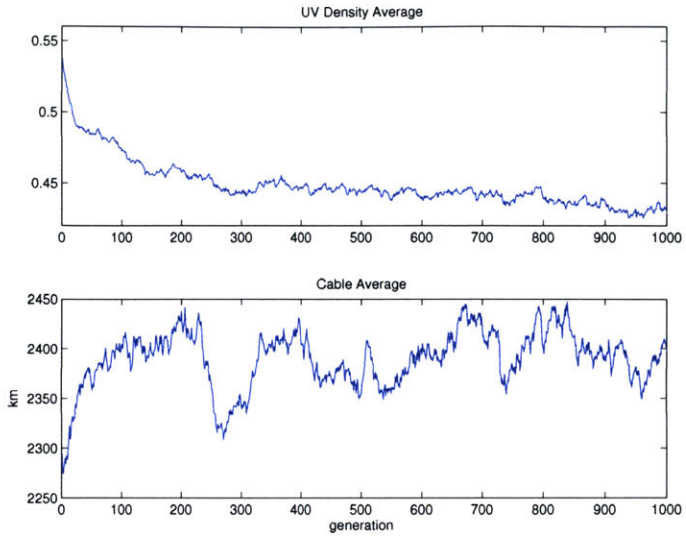


Figure A-38: 160-station, 20km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

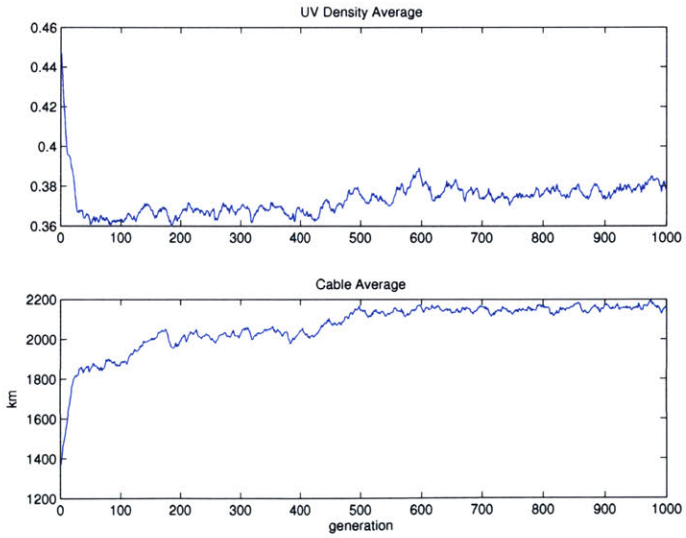


Figure A-39: 160-station, 40km *gridstep*, 10% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

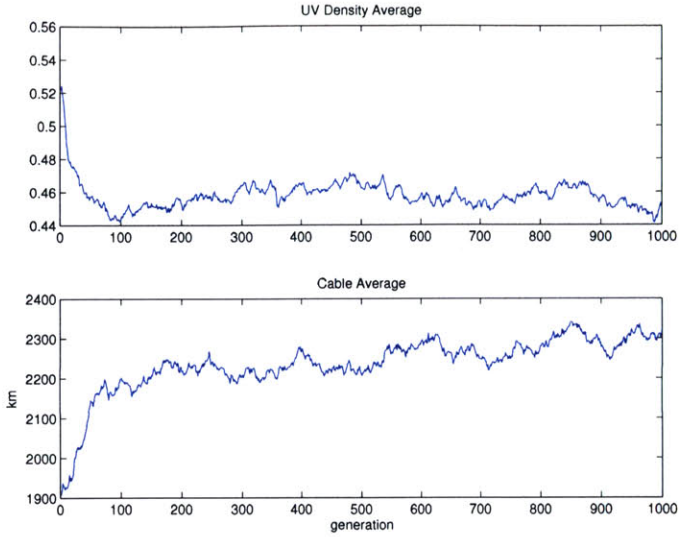


Figure A-40: 160-station, 40km *gridstep*, 30% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

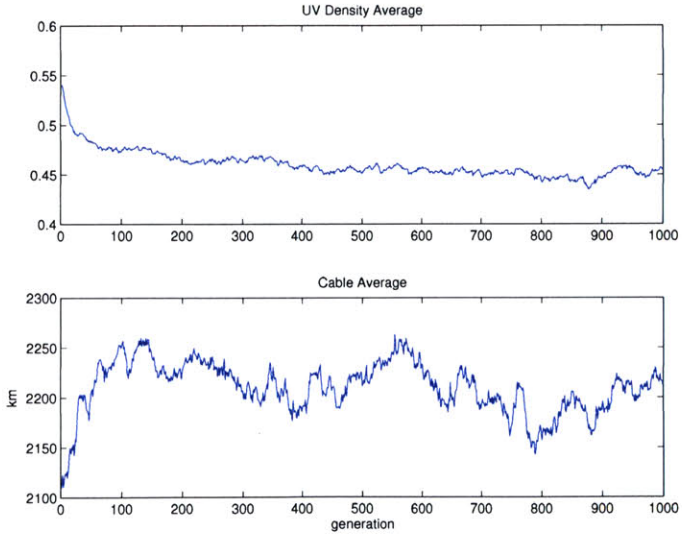


Figure A-41: 160-station, 40km *gridstep*, 50% badzones convergence information. The top plot shows the average performance metric value of the current generation as a function of generation. The bottom plot shows the average cable length value of the current generation as a function of generation.

A.4 Objective Spaces for Site Constraint Results

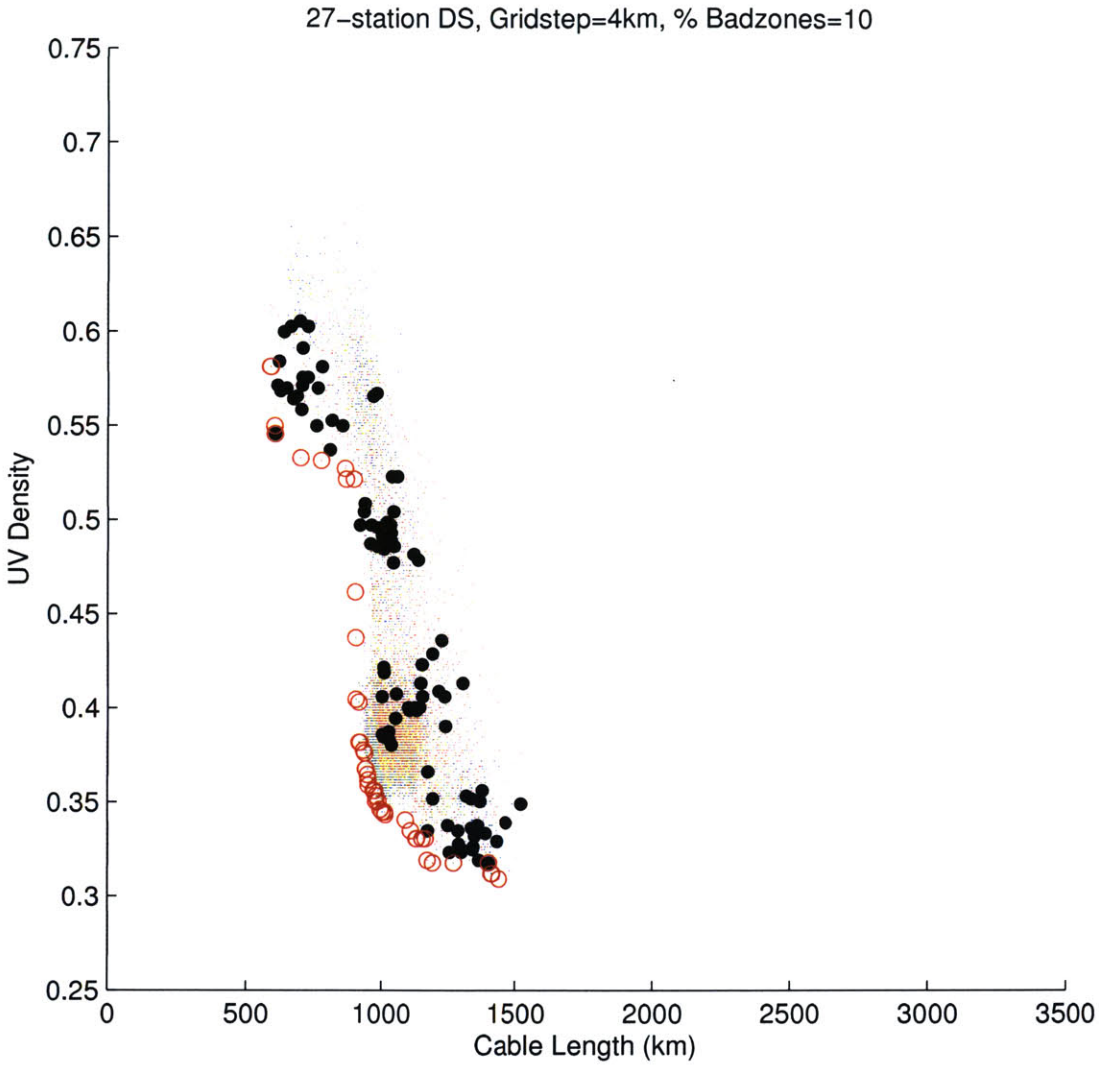


Figure A-42: 27-station, 4km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

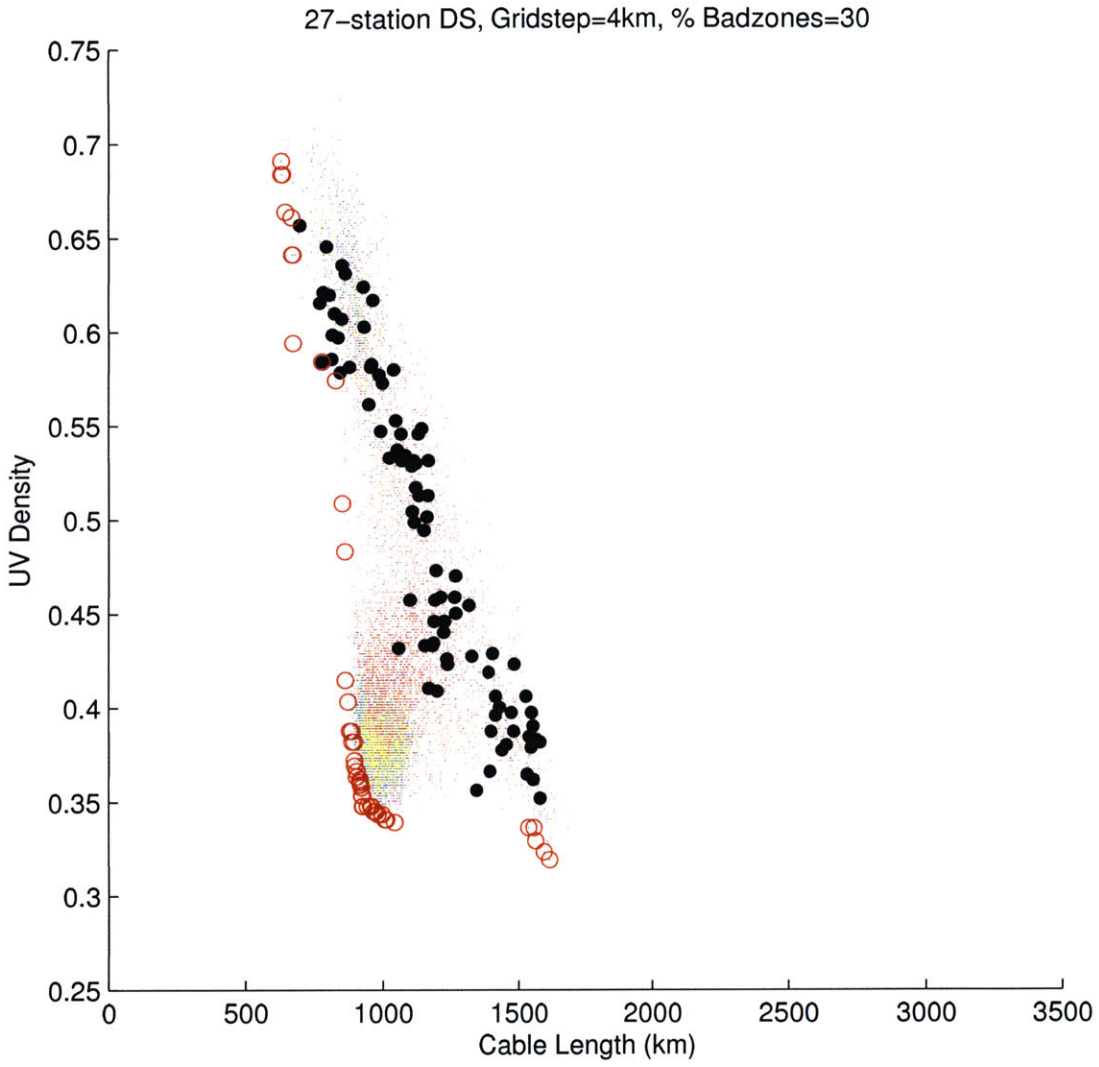


Figure A-43: 27-station, 4km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

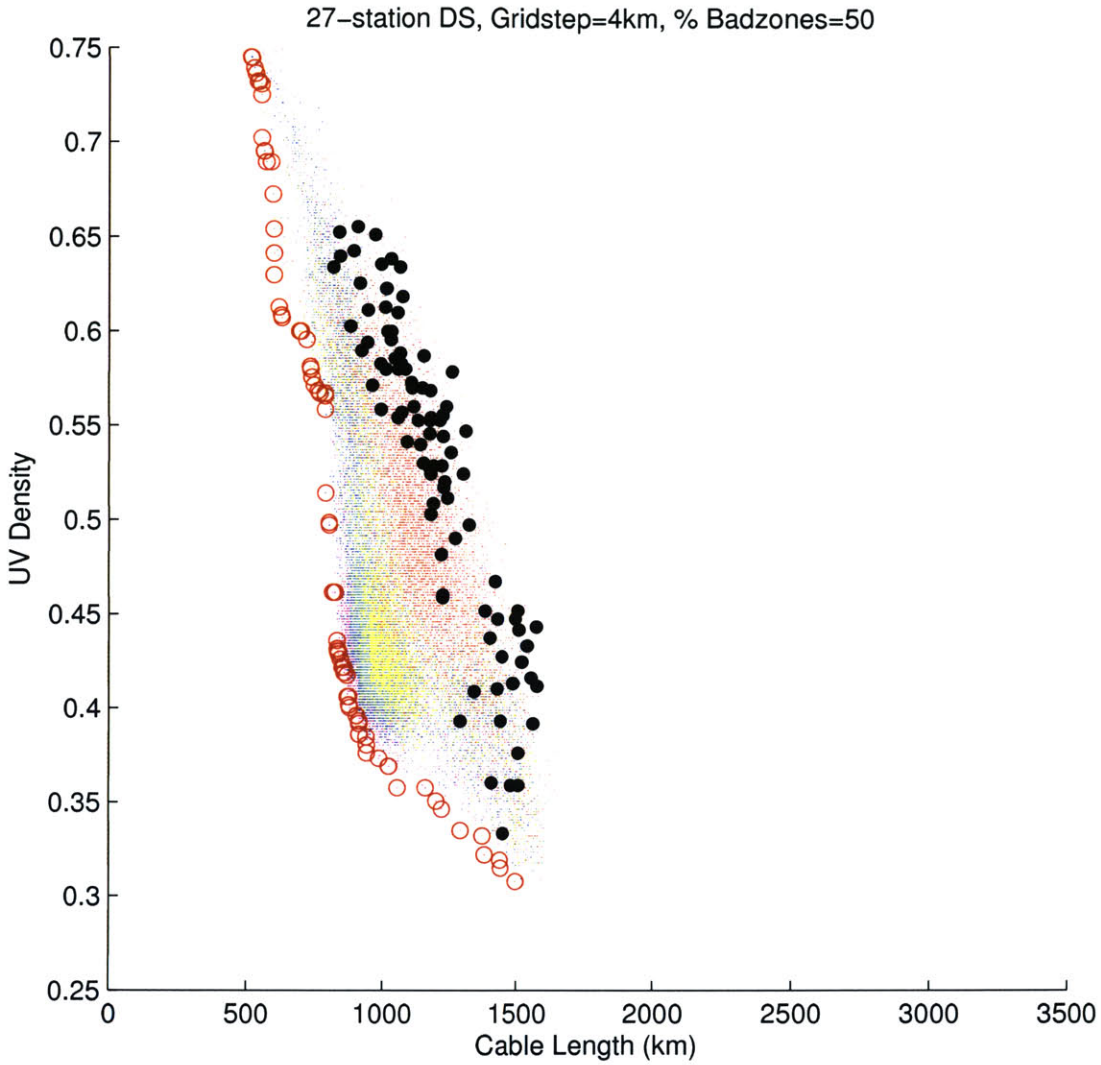


Figure A-44: 27-station, 4km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

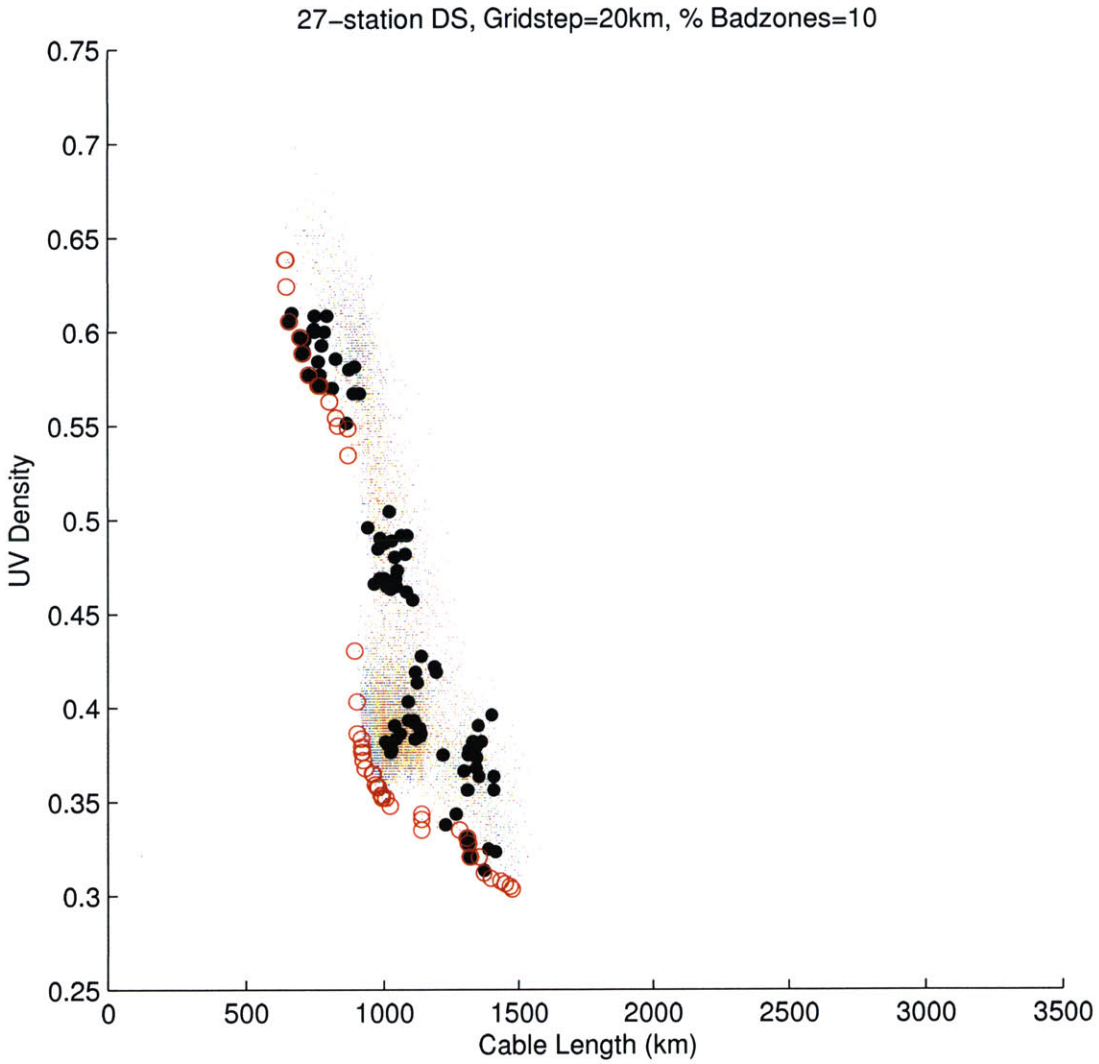


Figure A-45: 27-station, 20km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

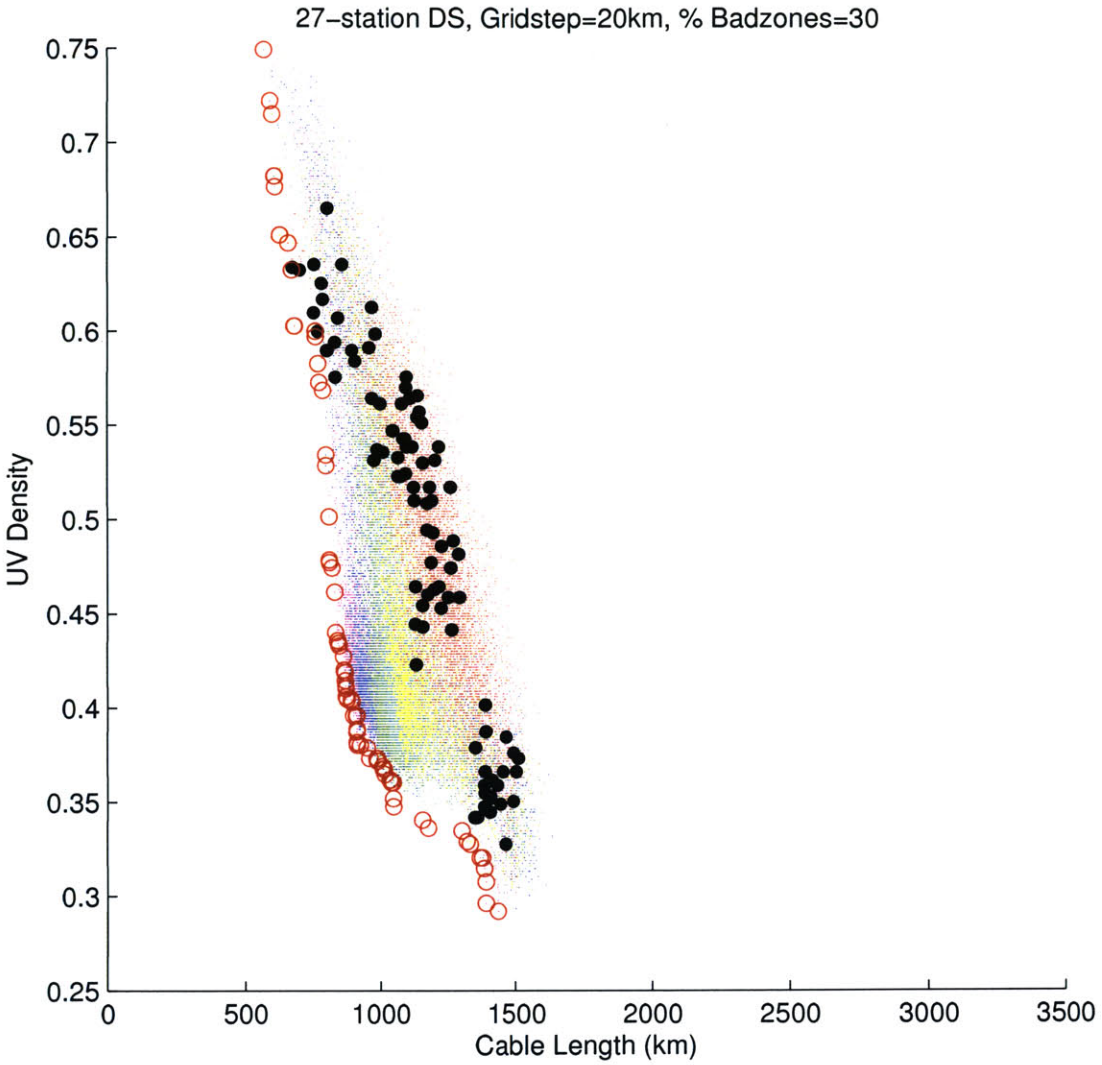


Figure A-46: 27-station, 20km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

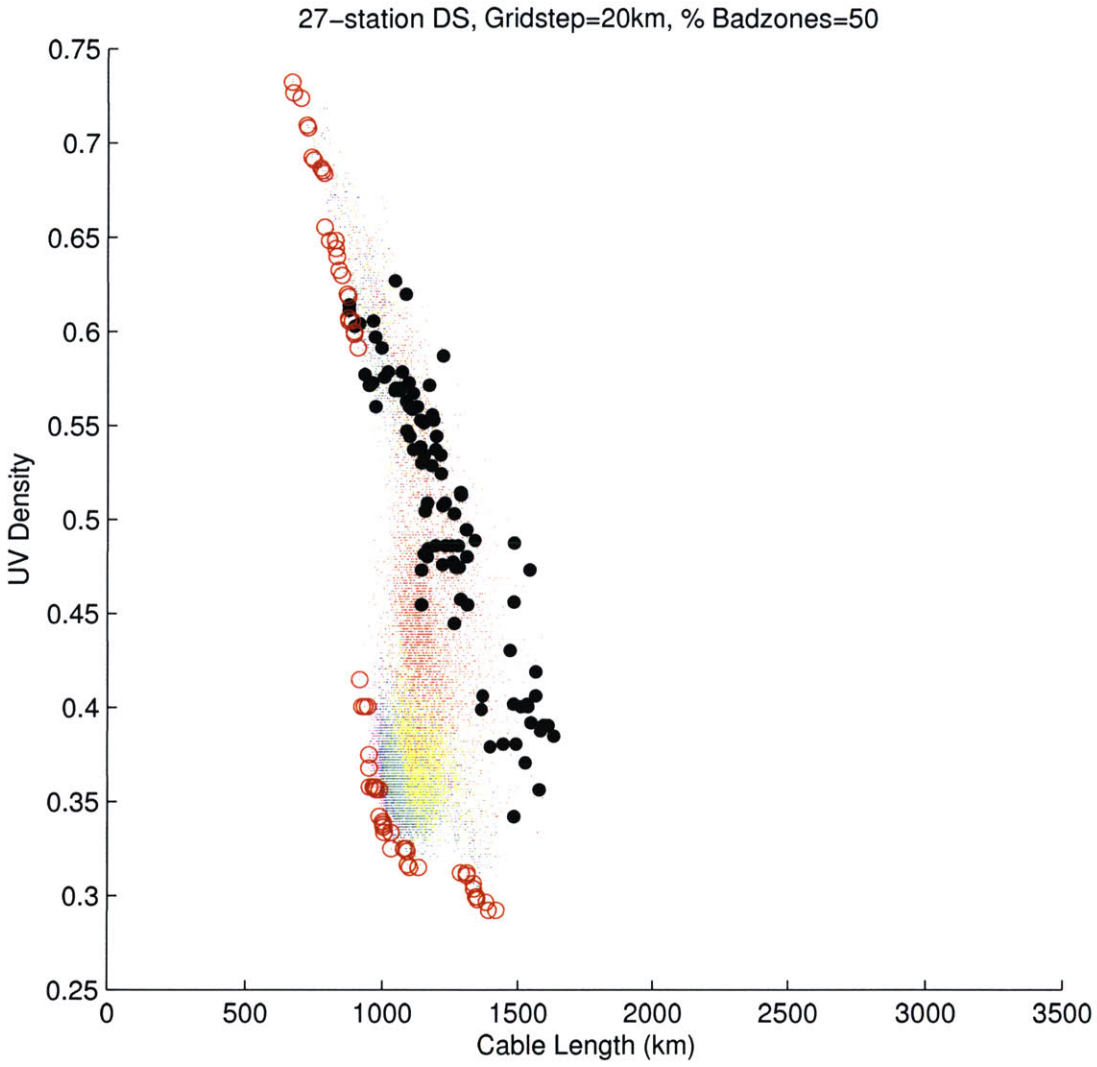


Figure A-47: 27-station, 20km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

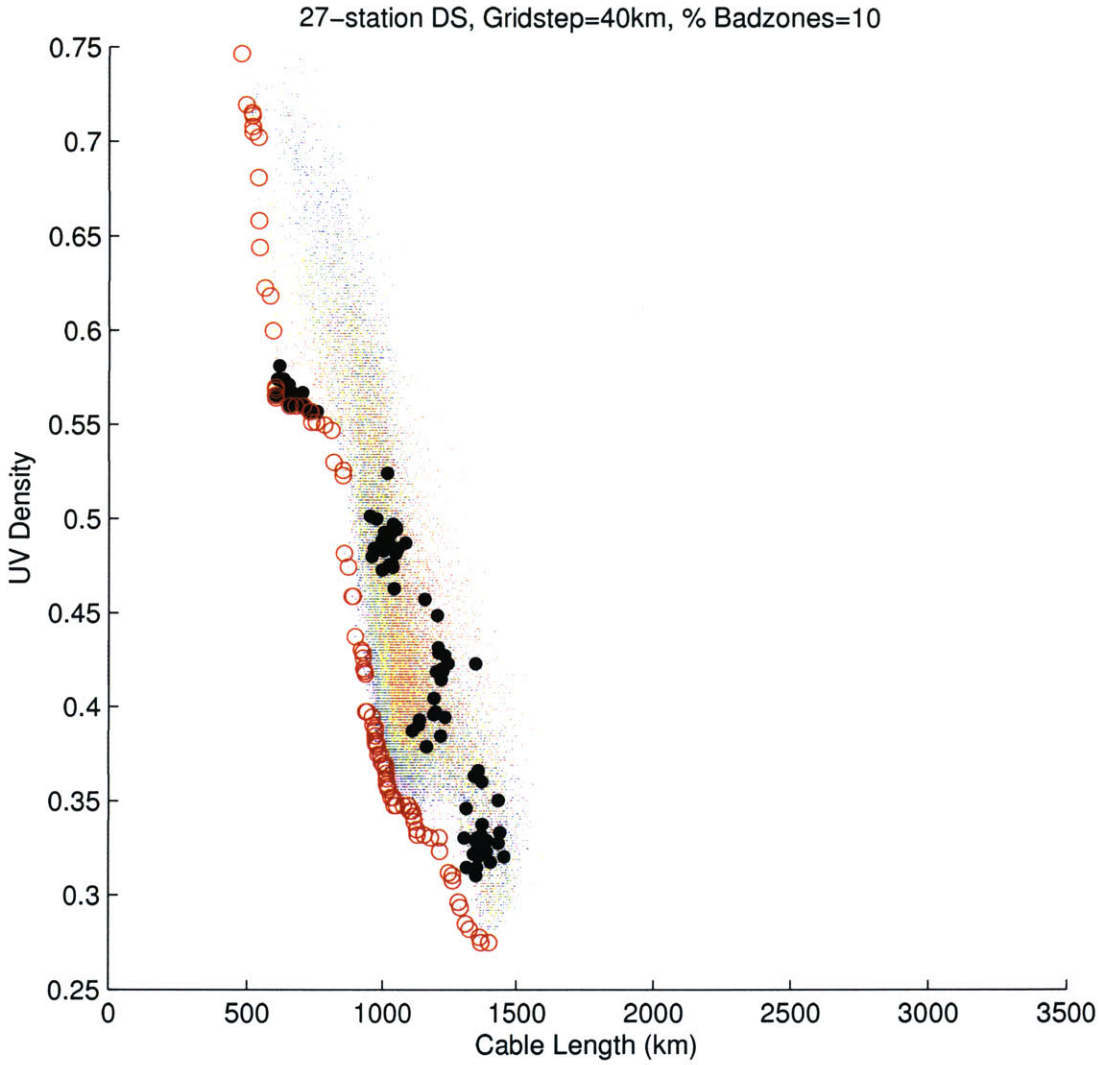


Figure A-48: 27-station, 40km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

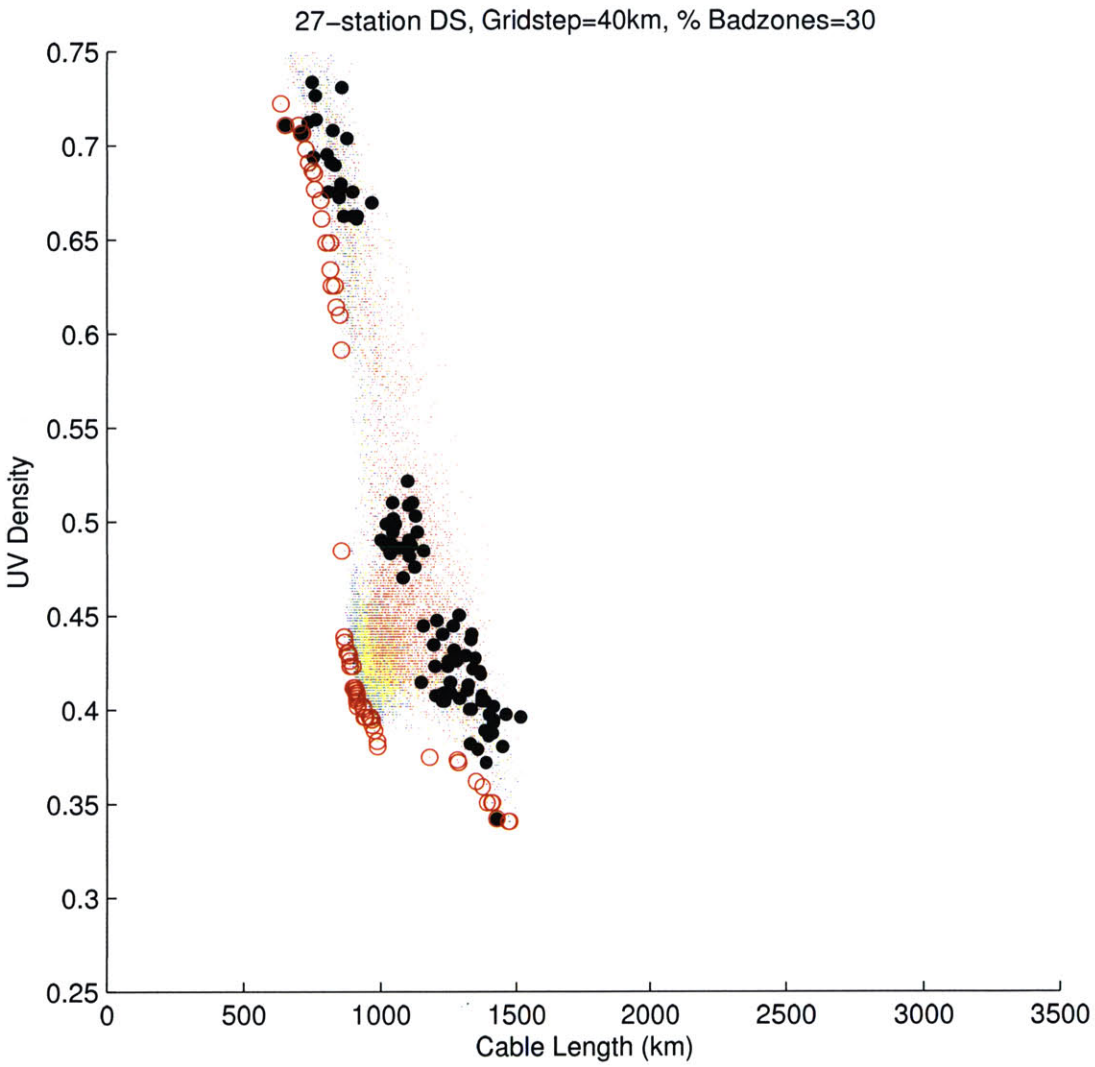


Figure A-49: 27-station, 40km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

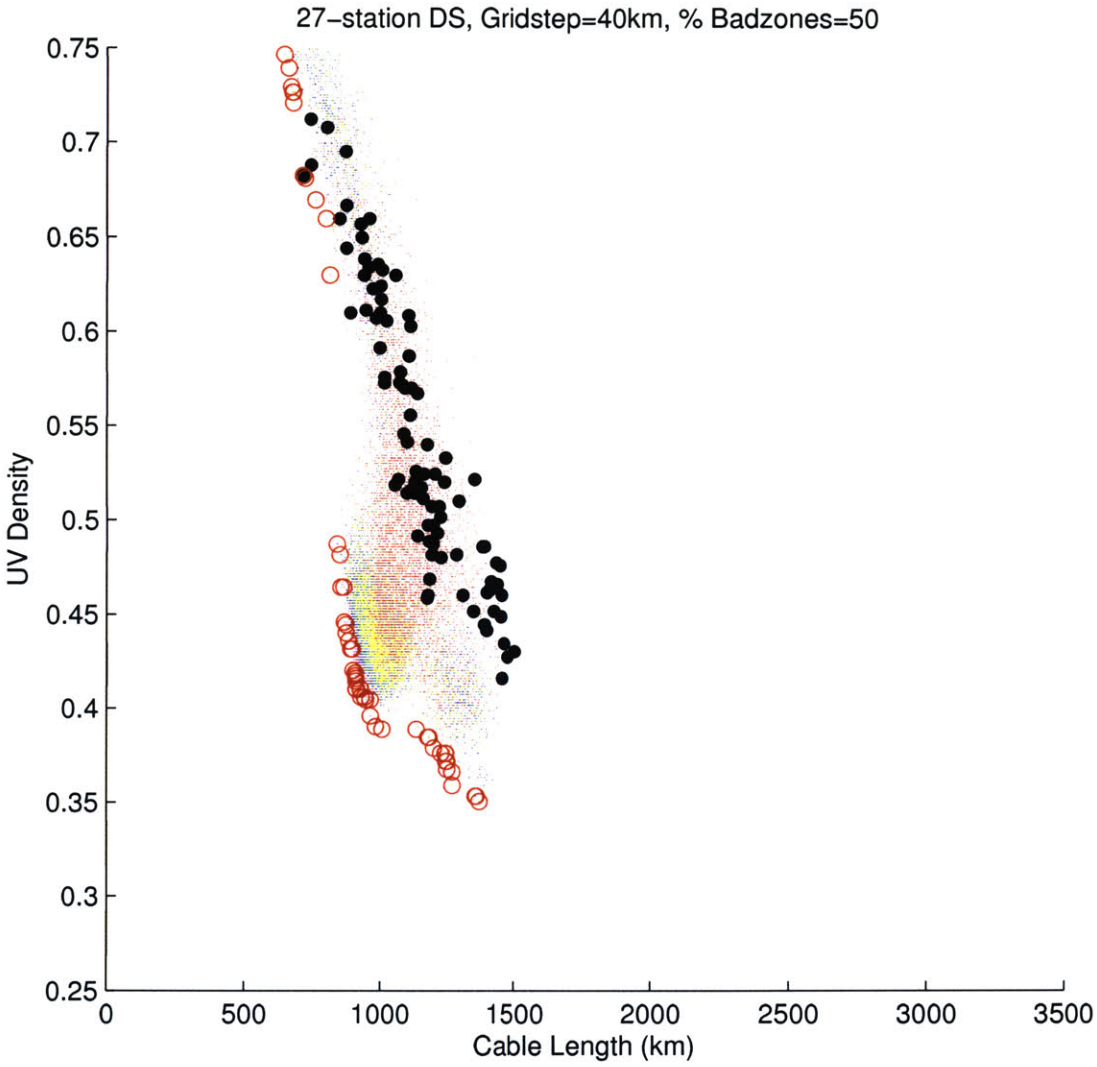


Figure A-50: 27-station, 40km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

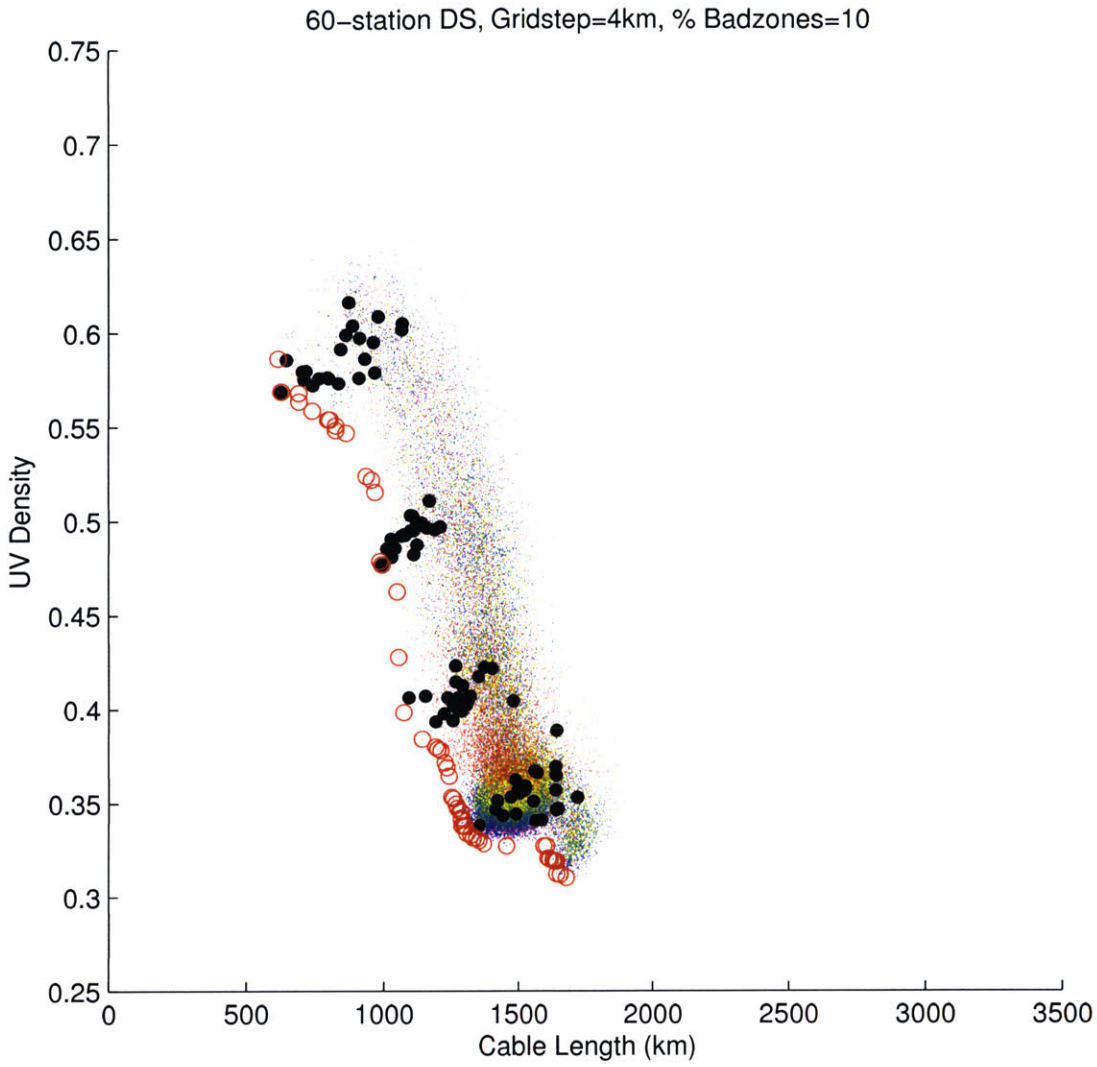


Figure A-51: 60-station, 4km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

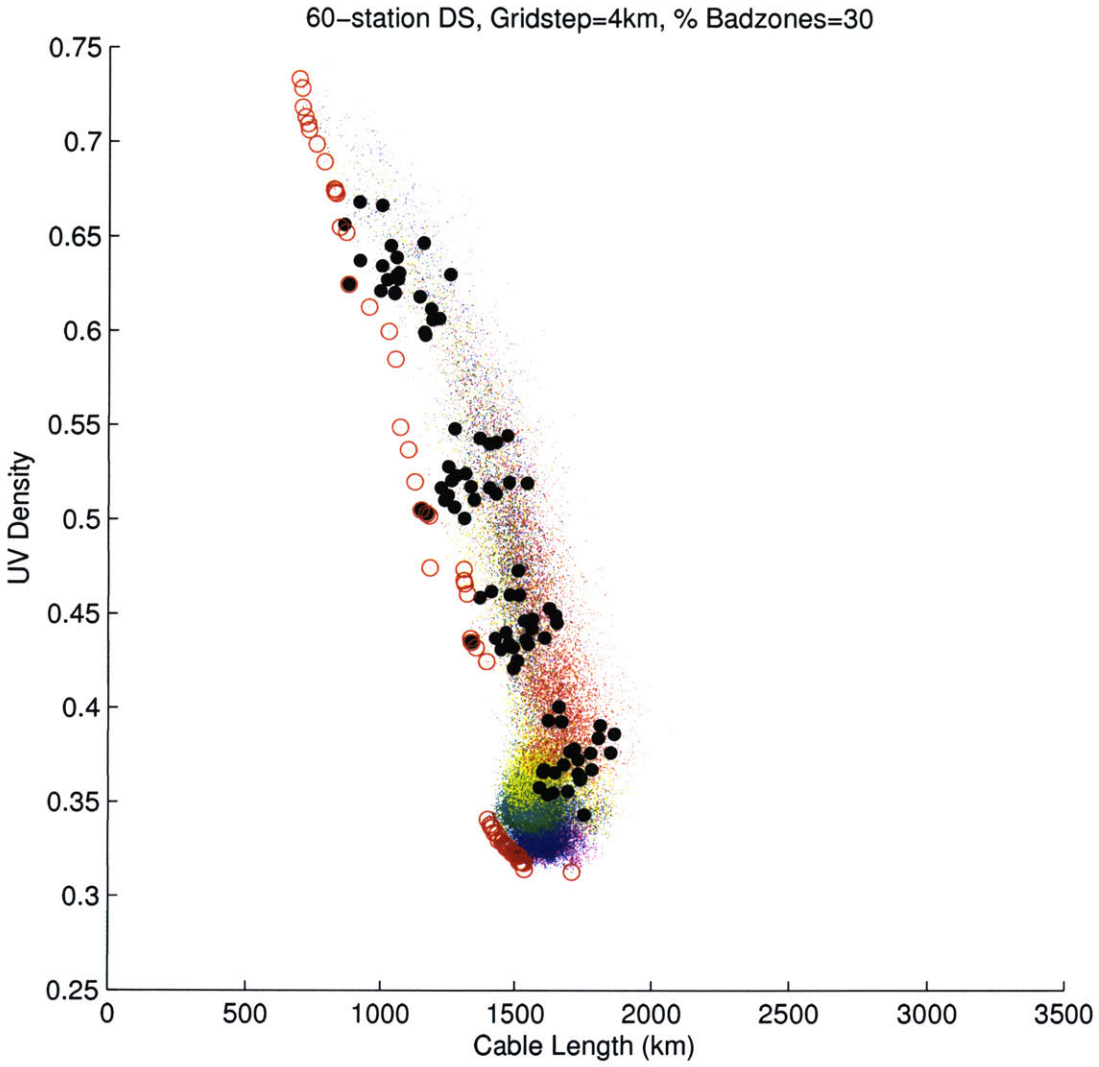


Figure A-52: 60-station, 4km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

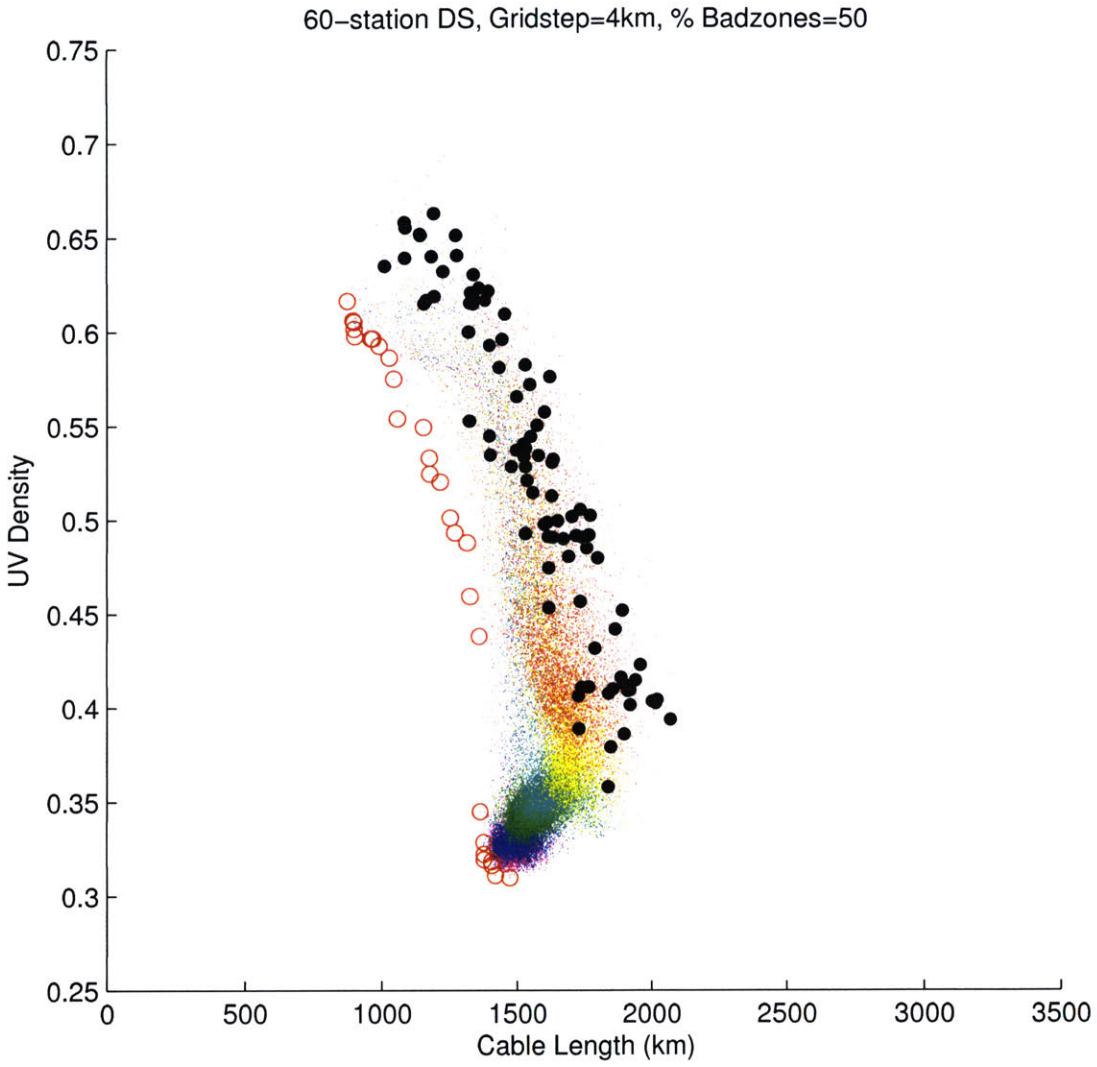


Figure A-53: 60-station, 4km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

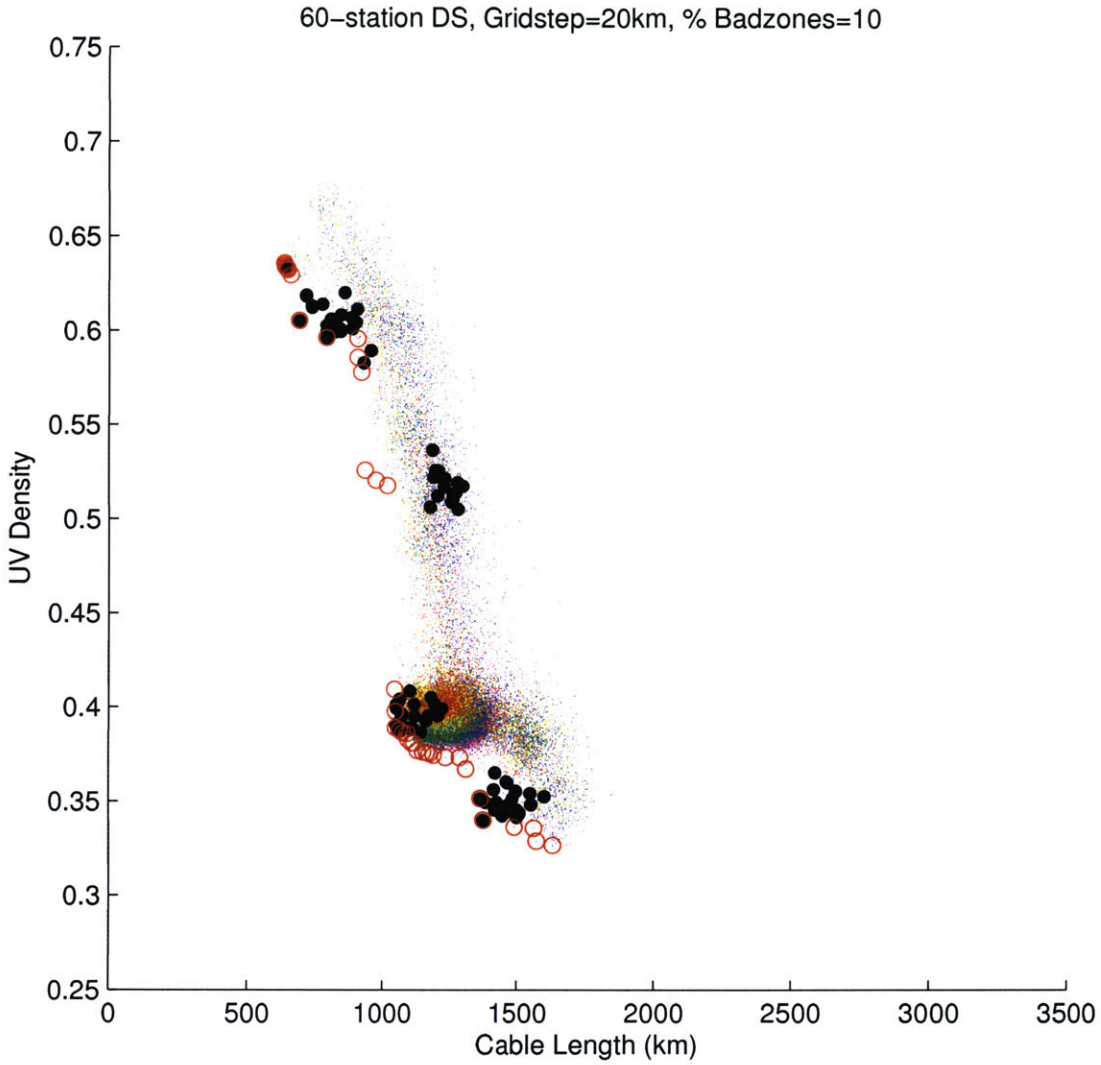


Figure A-54: 60-station, 20km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

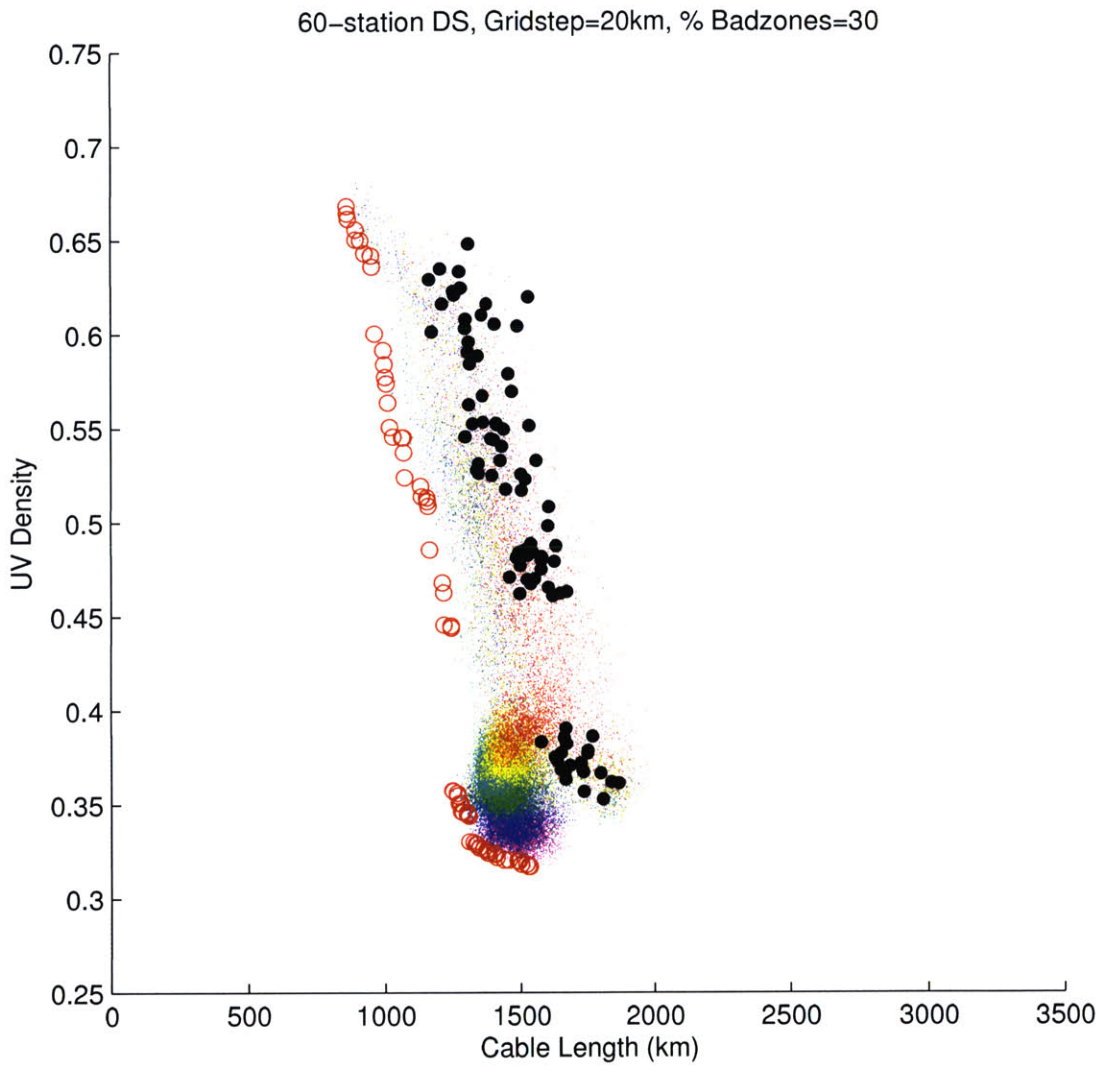


Figure A-55: 60-station, 20km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

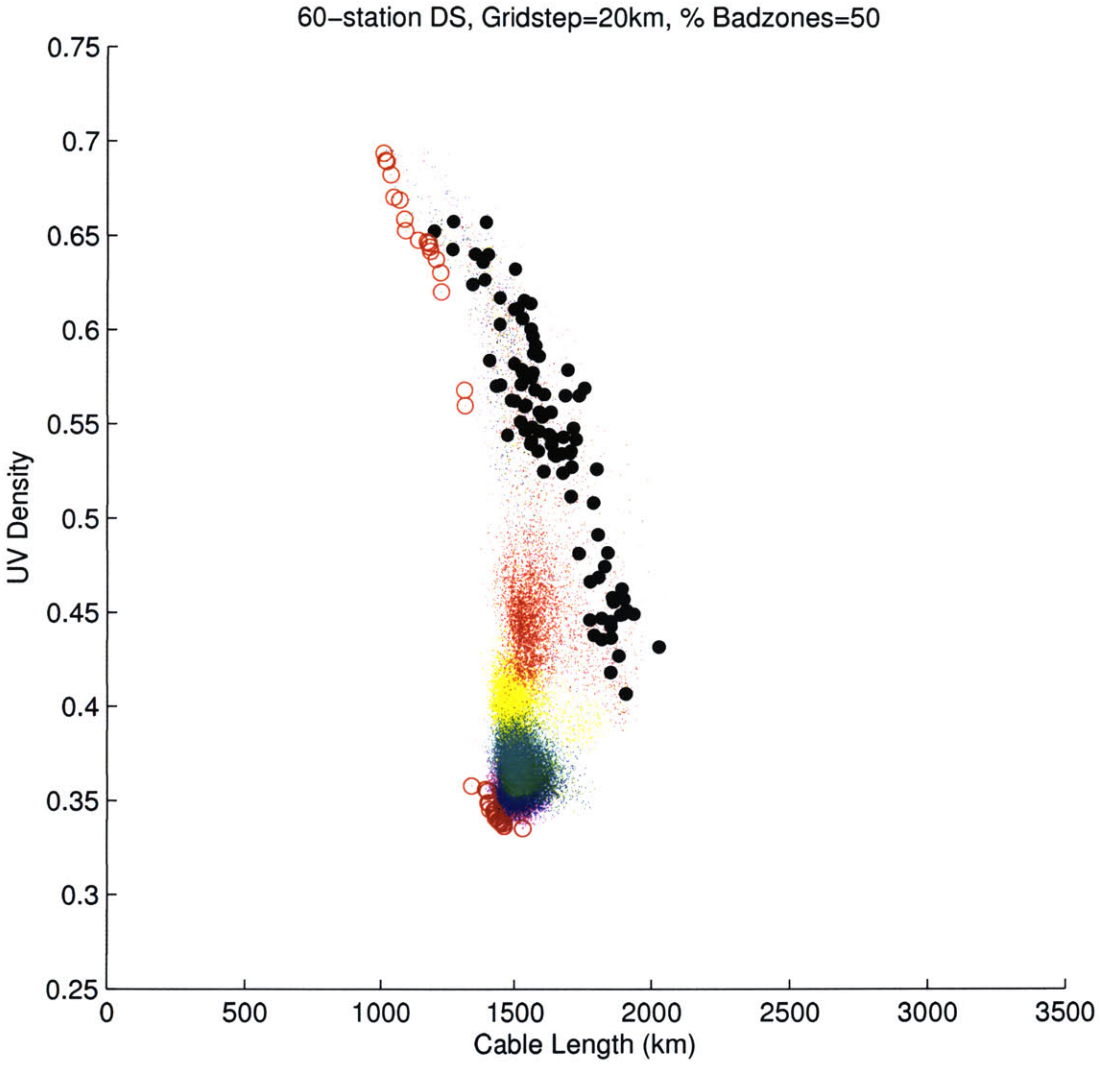


Figure A-56: 60-station, 20km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

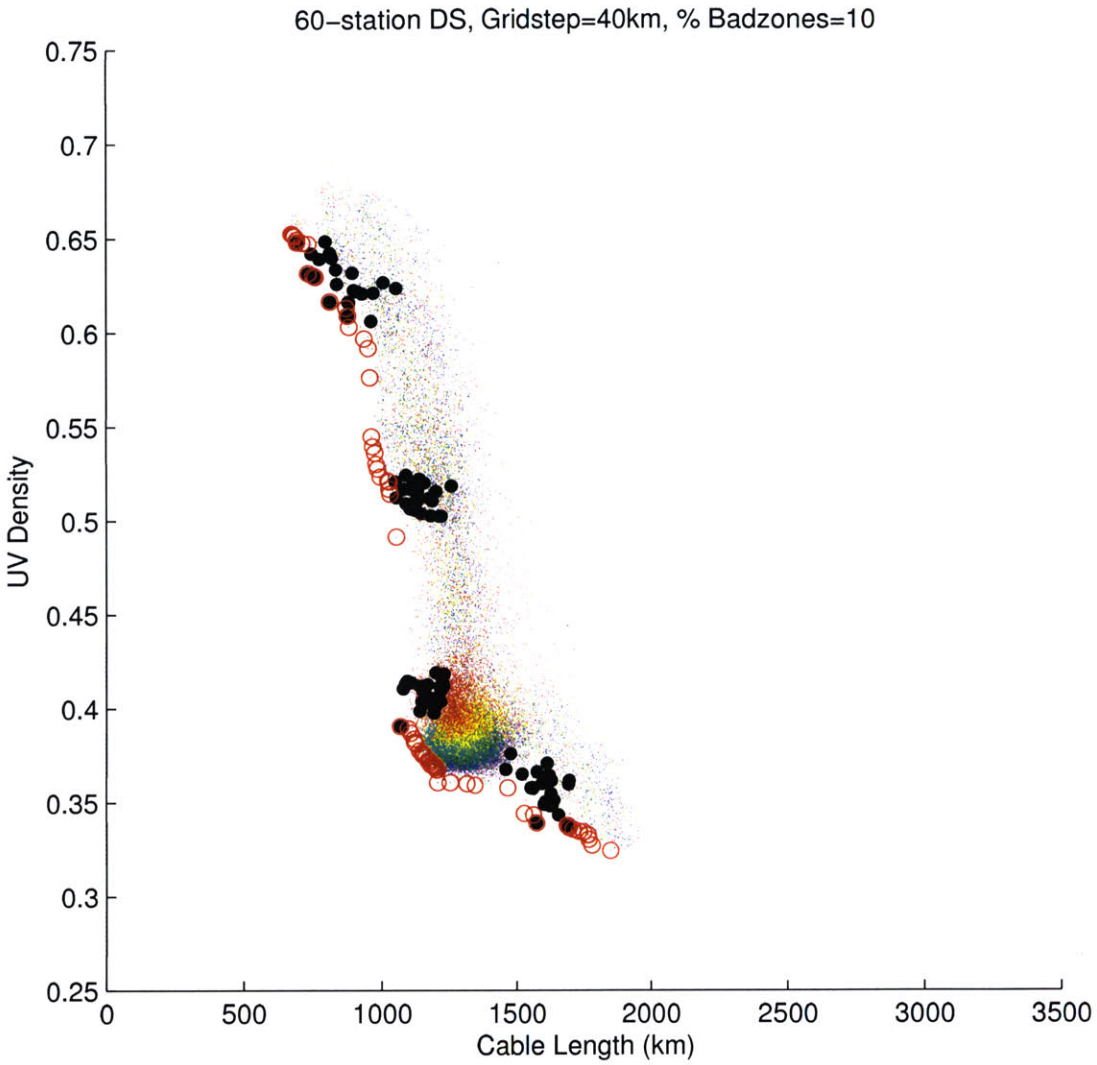


Figure A-57: 60-station, 40km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

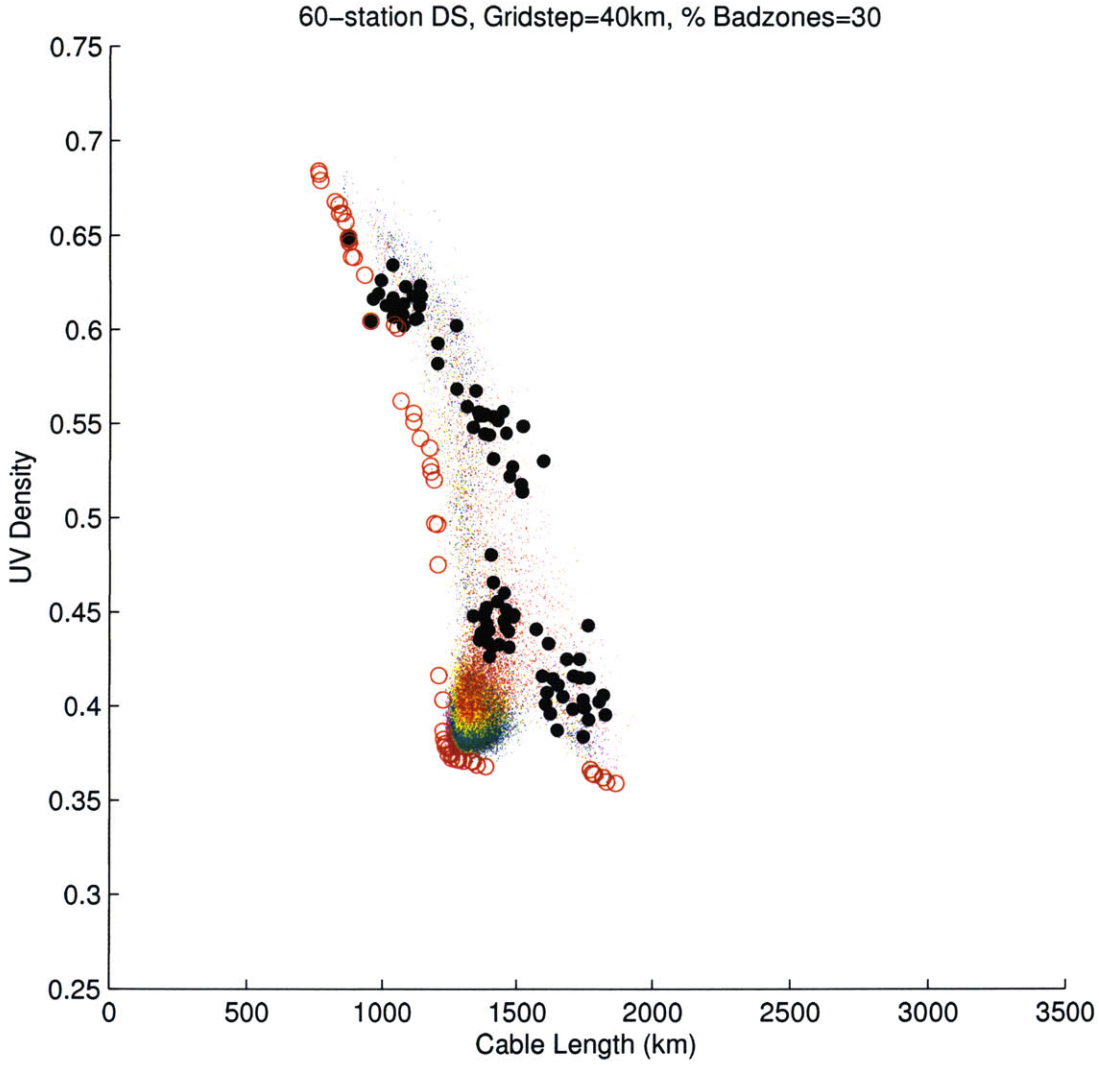


Figure A-58: 60-station, 40km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

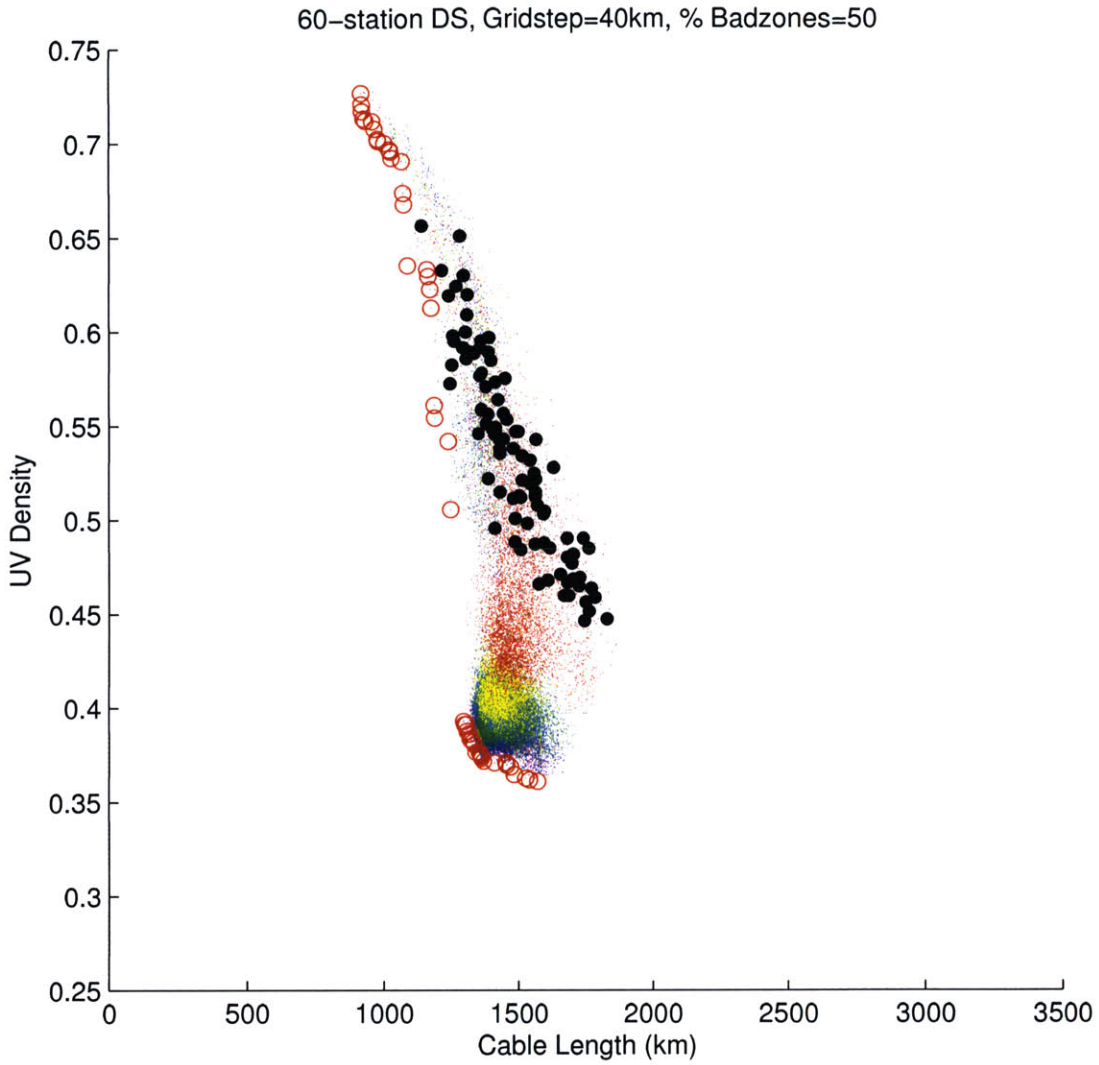


Figure A-59: 60-station, 40km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

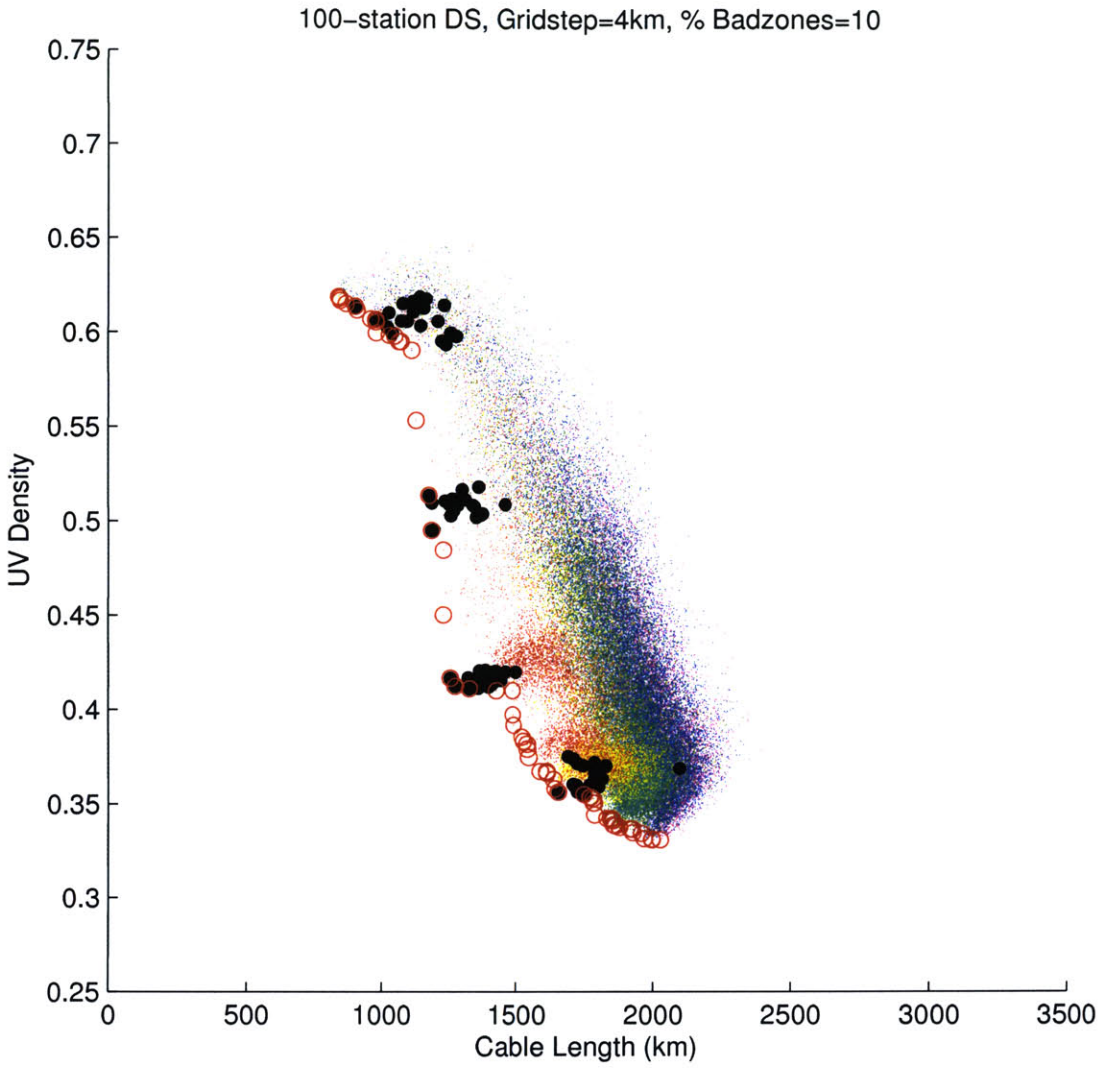


Figure A-60: 100-station, 4km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

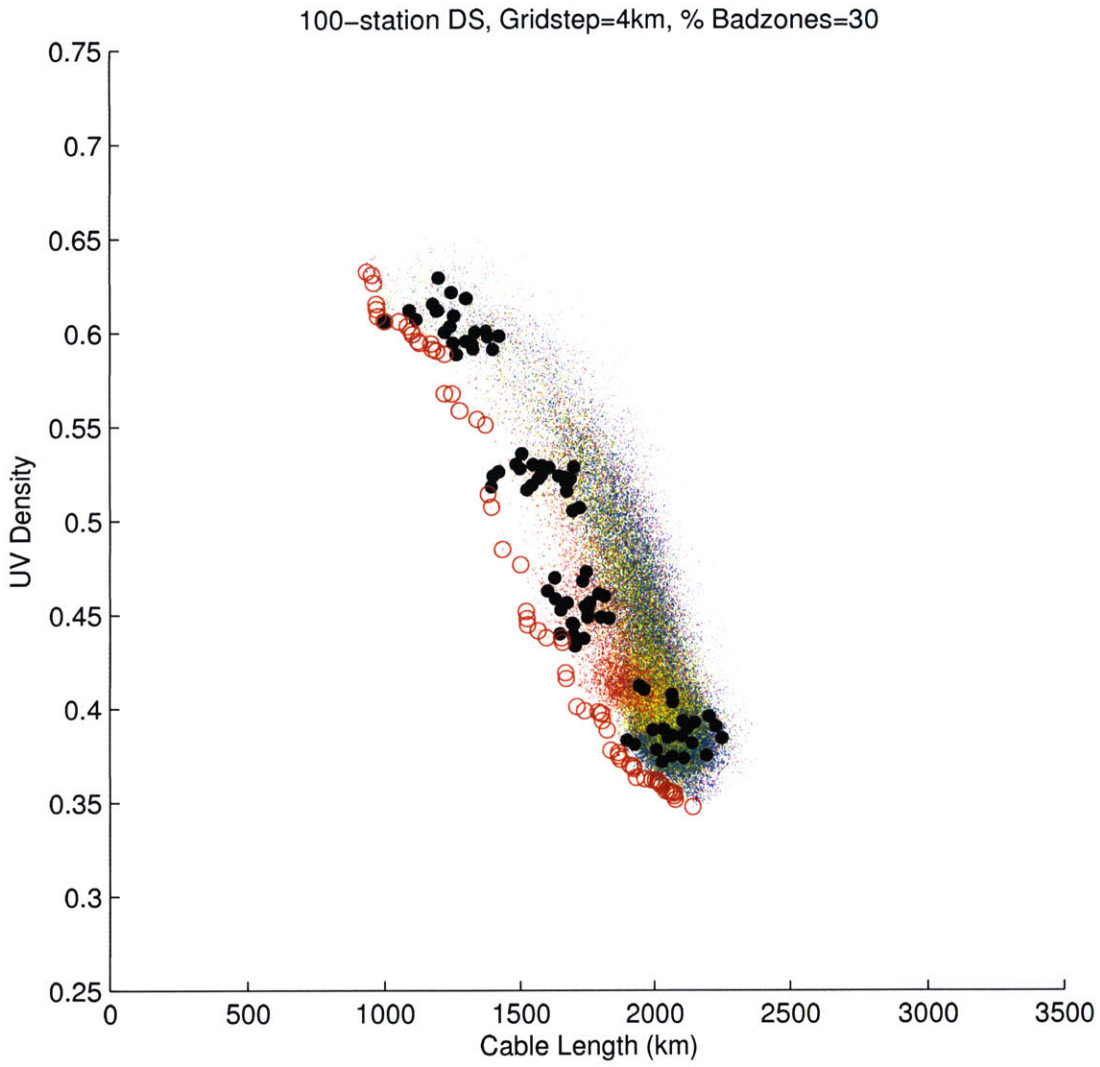


Figure A-61: 100-station, 4km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

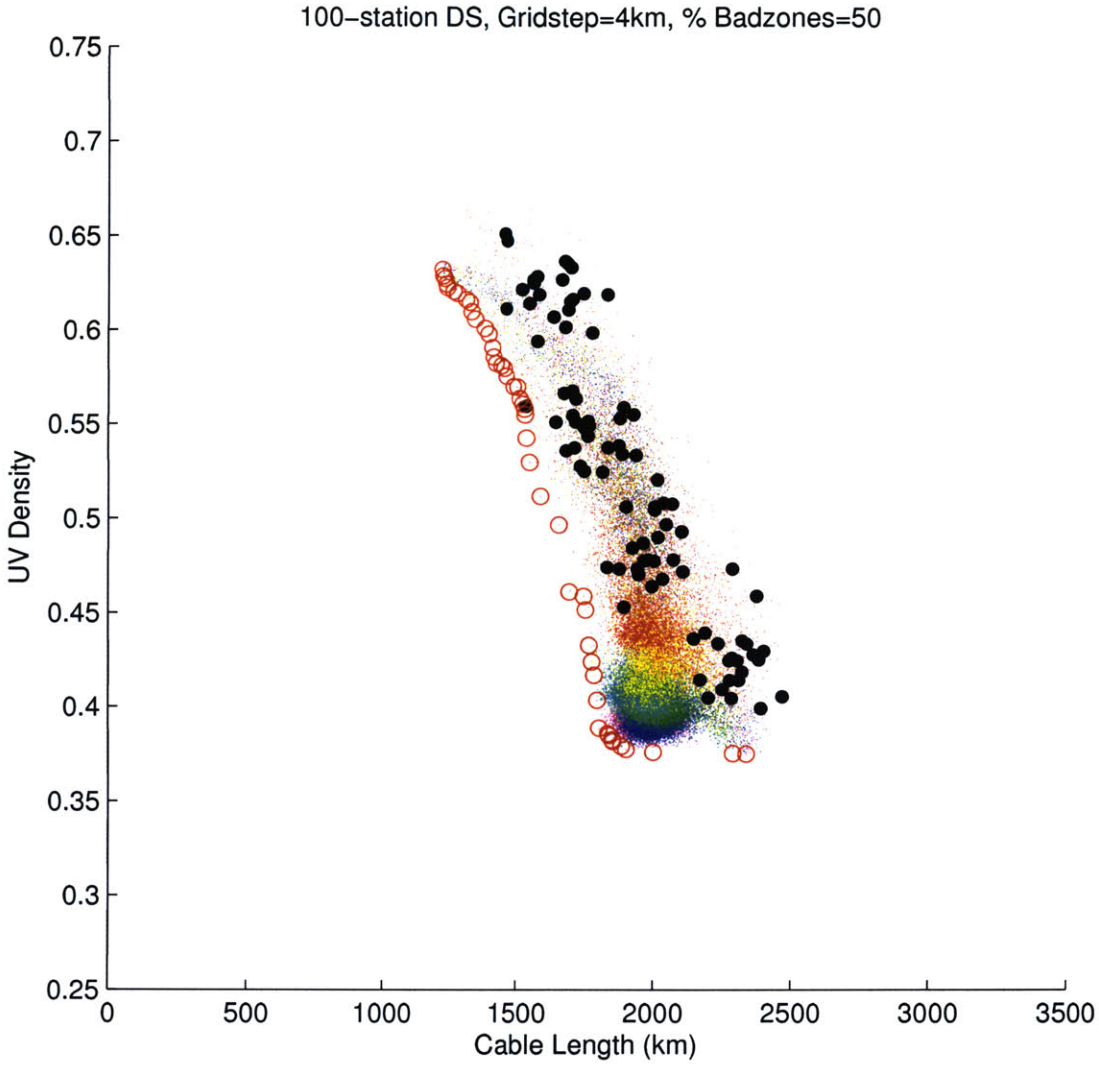


Figure A-62: 100-station, 4km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

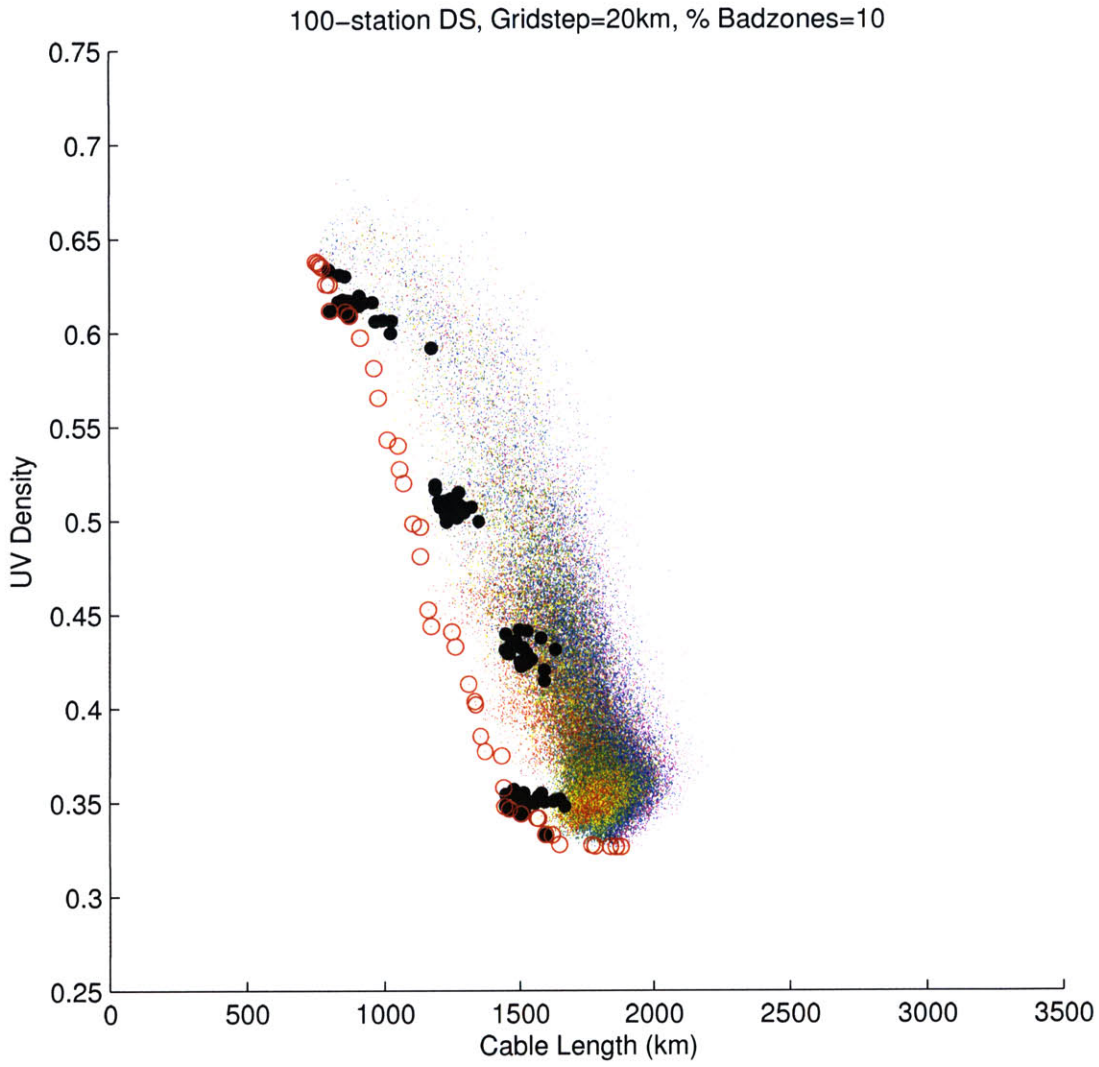


Figure A-63: 100-station, 20km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

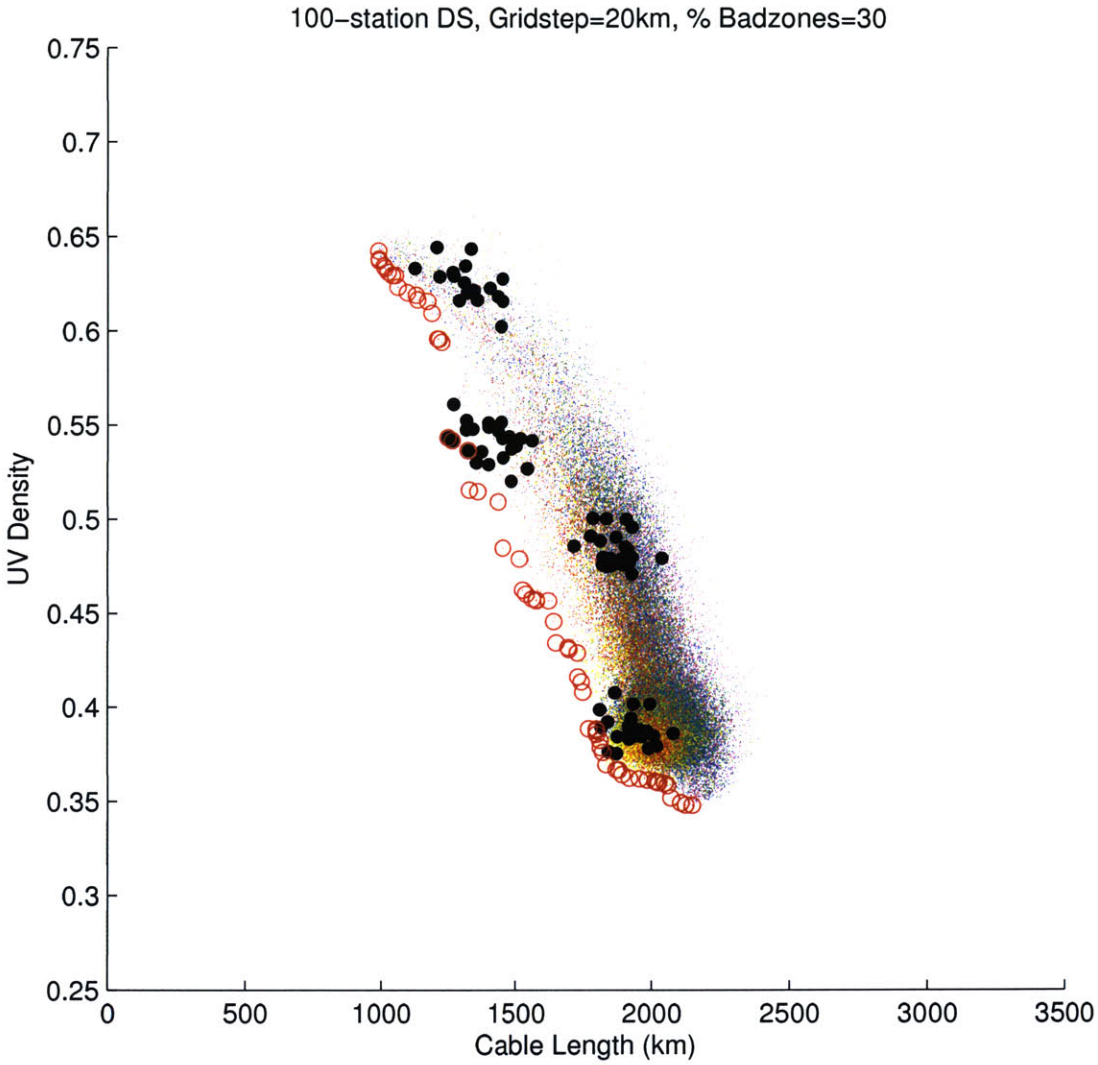


Figure A-64: 100-station, 20km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

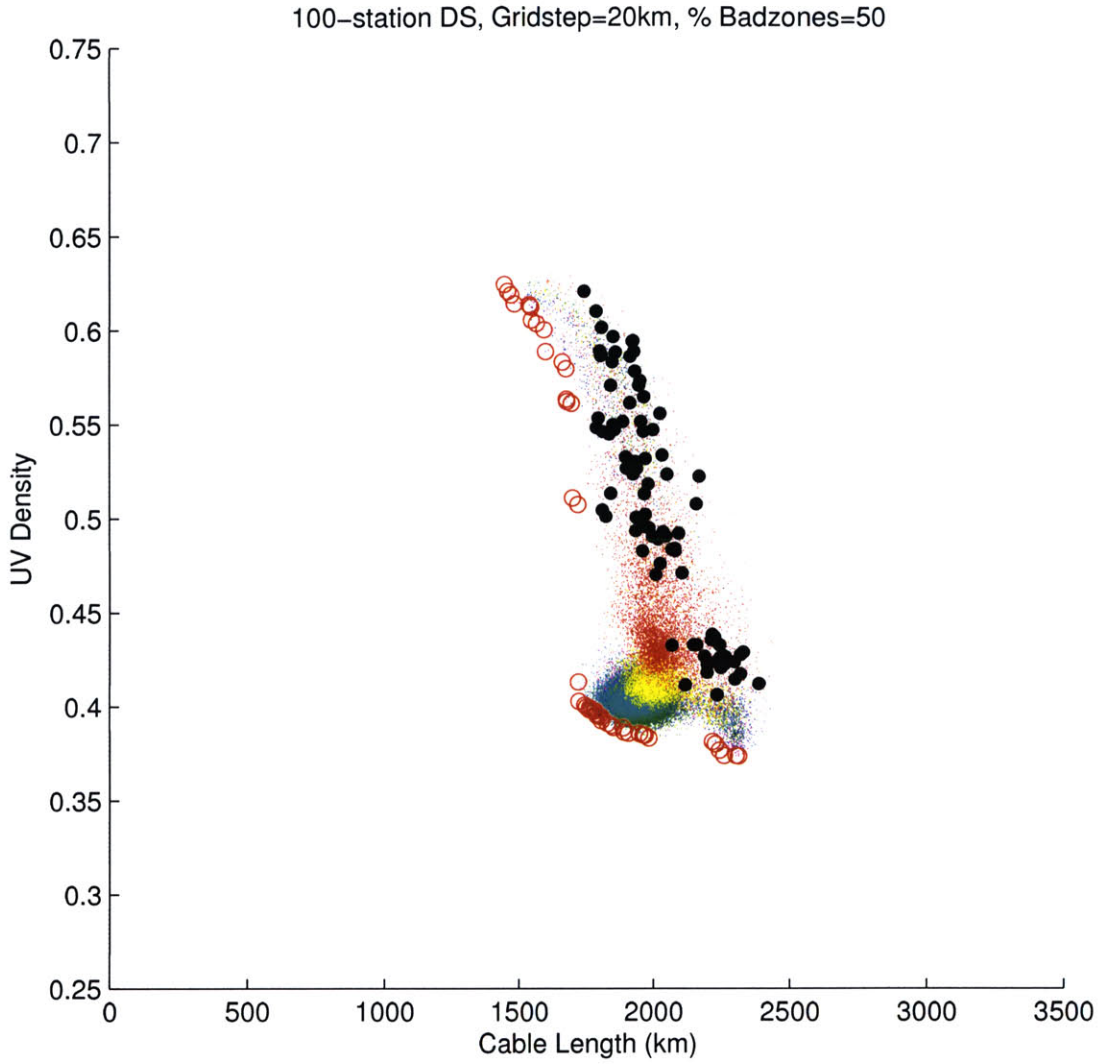


Figure A-65: 100-station, 20km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

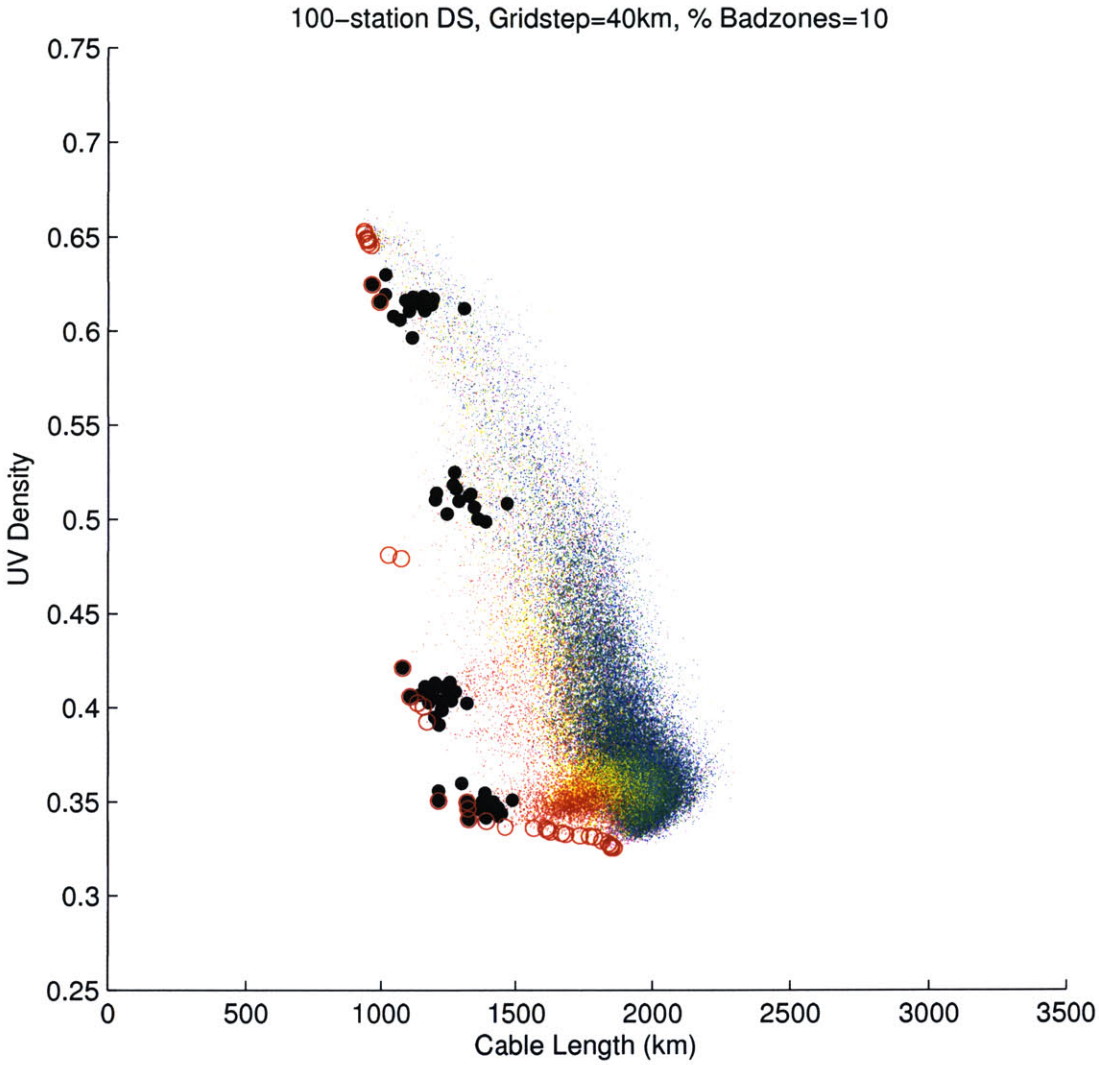


Figure A-66: 100-station, 40km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

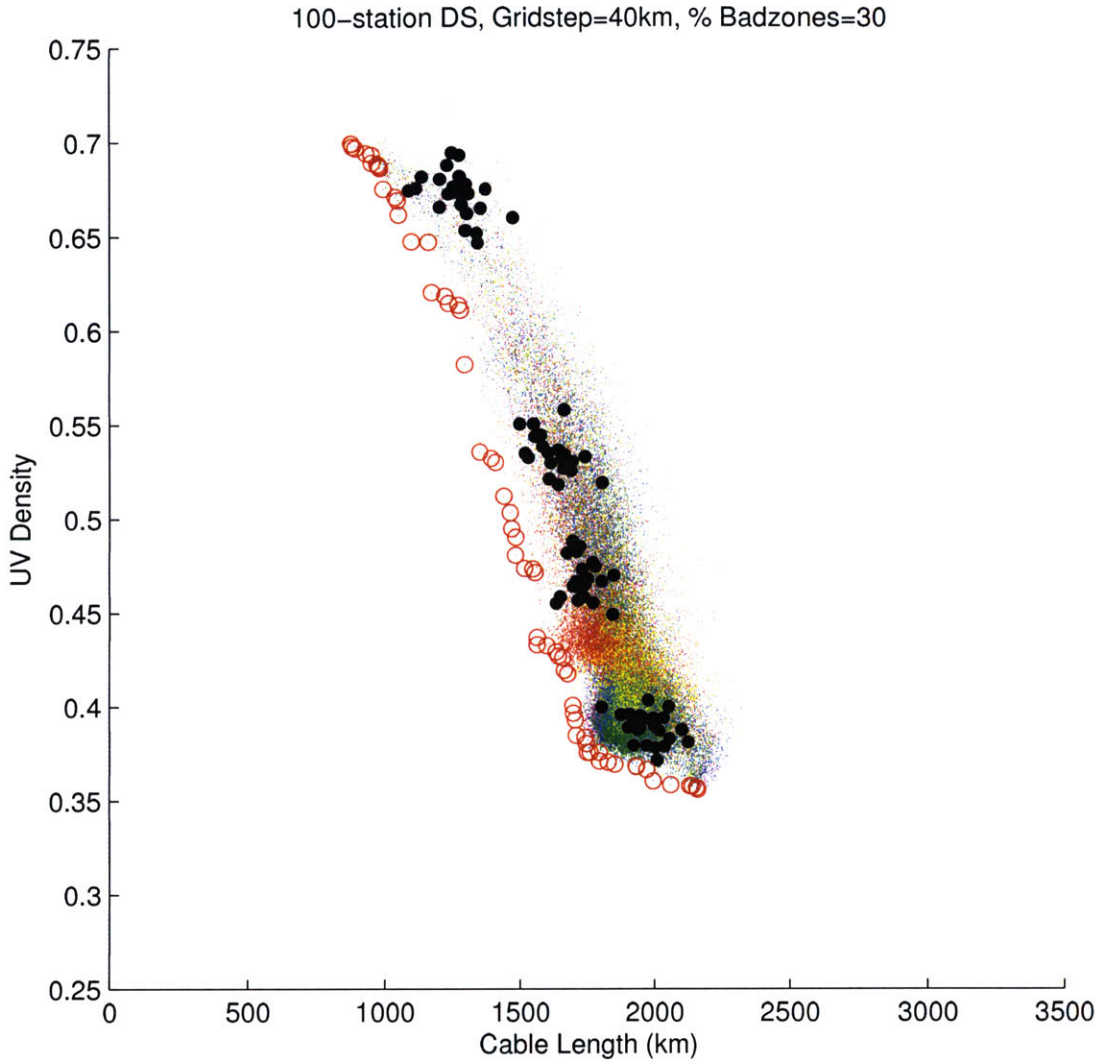


Figure A-67: 100-station, 40km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

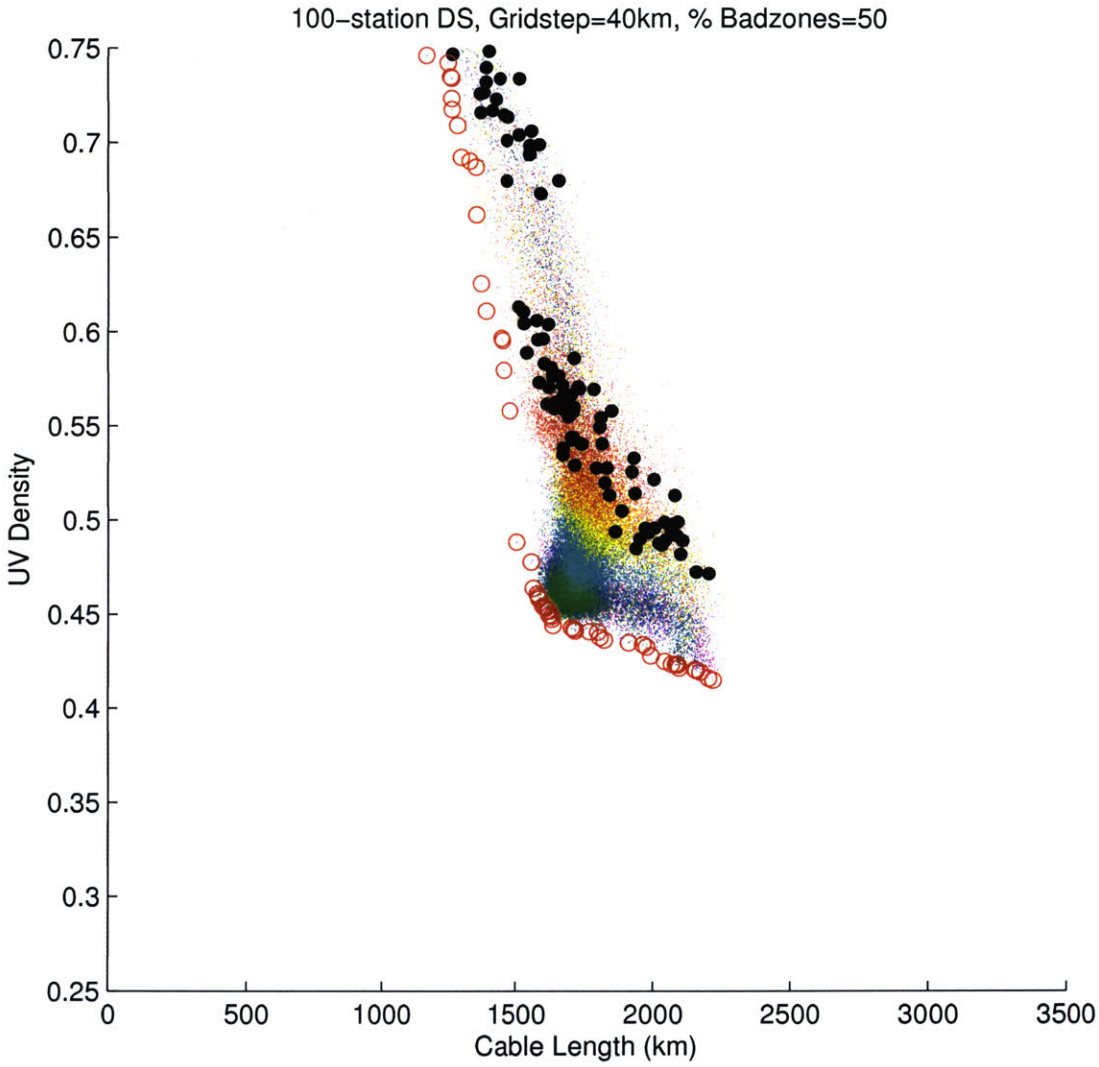


Figure A-68: 100-station, 40km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

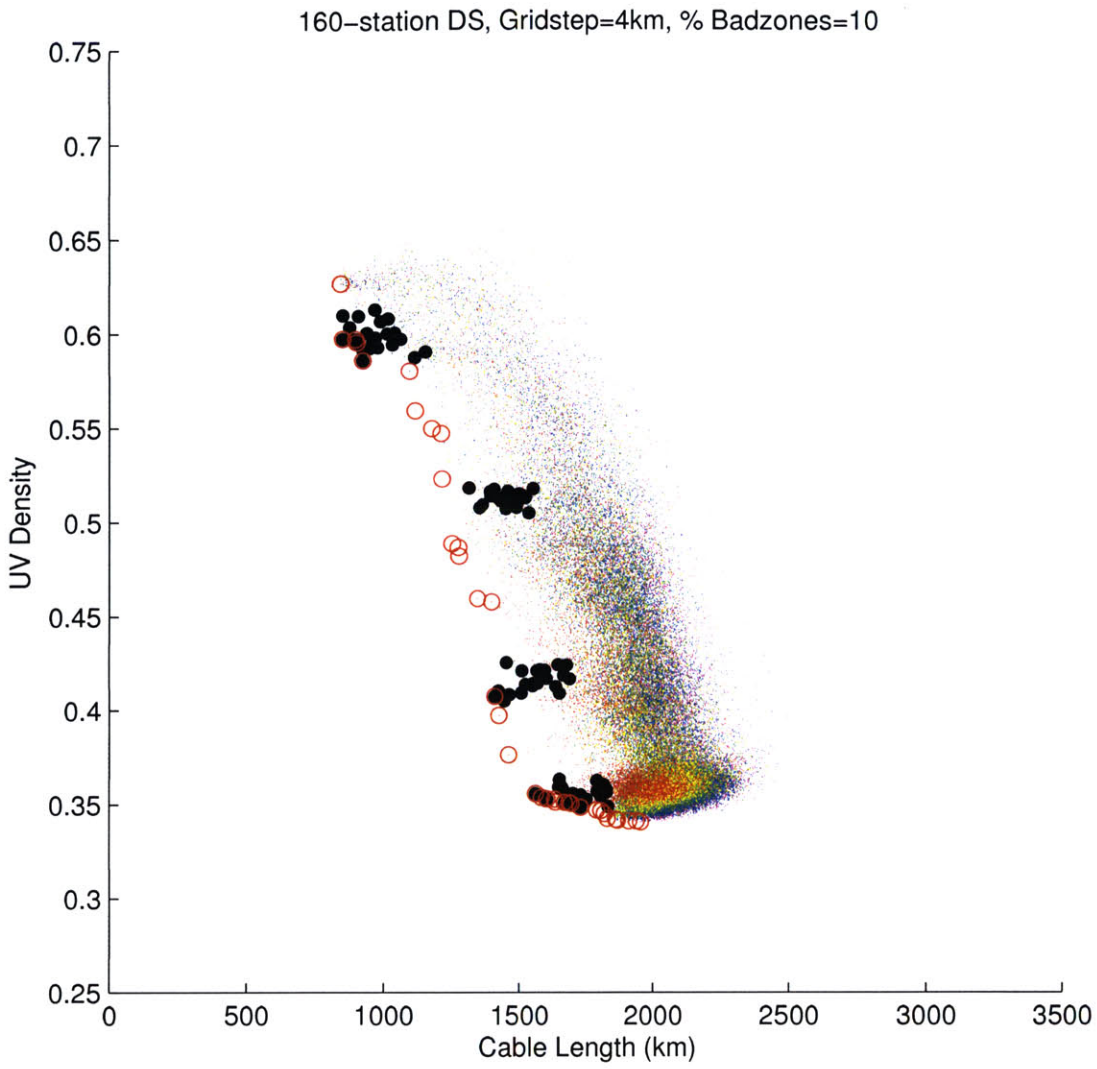


Figure A-69: 160-station, 4km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

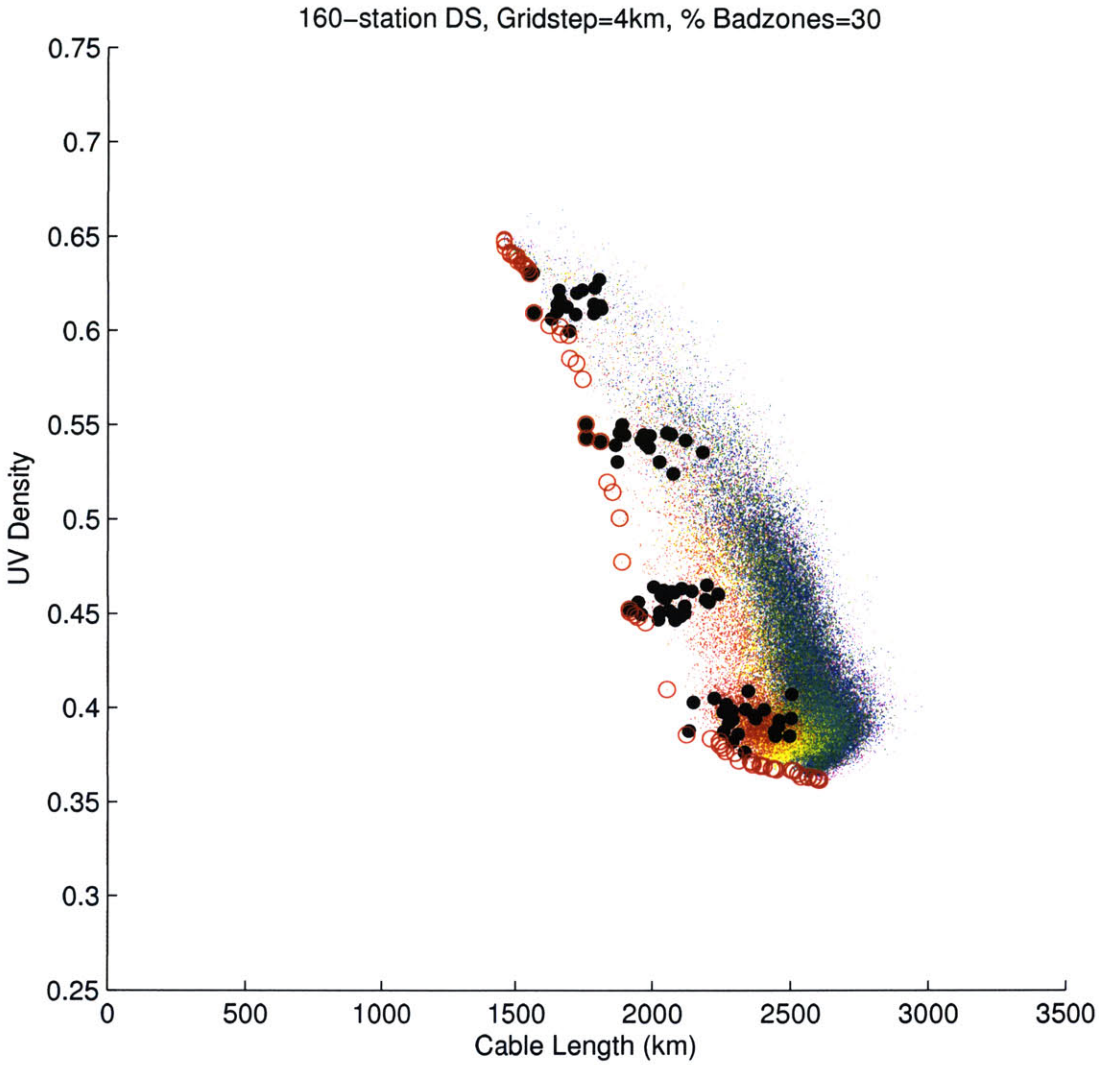


Figure A-70: 160-station, 4km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

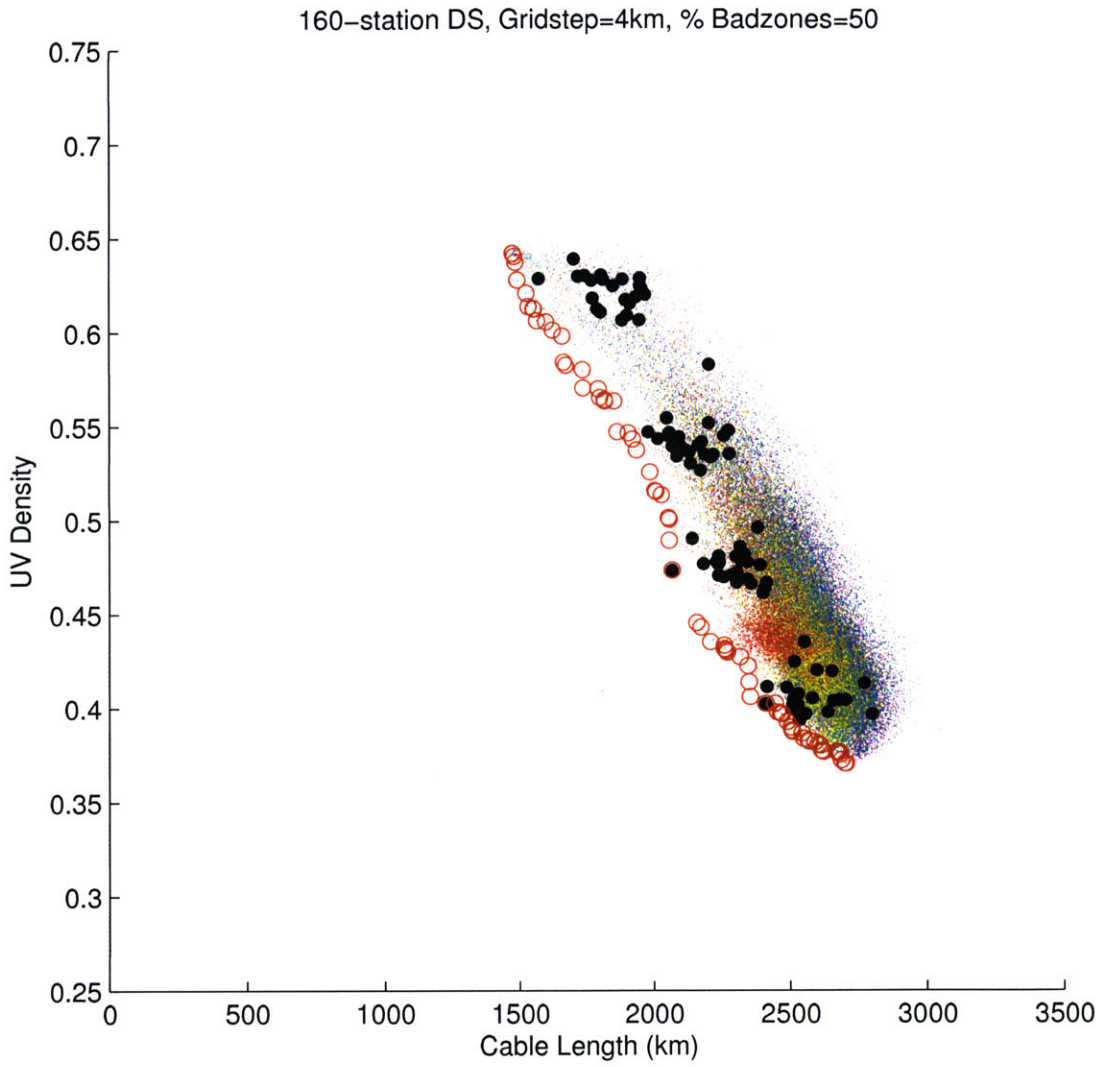


Figure A-71: 160-station, 4km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

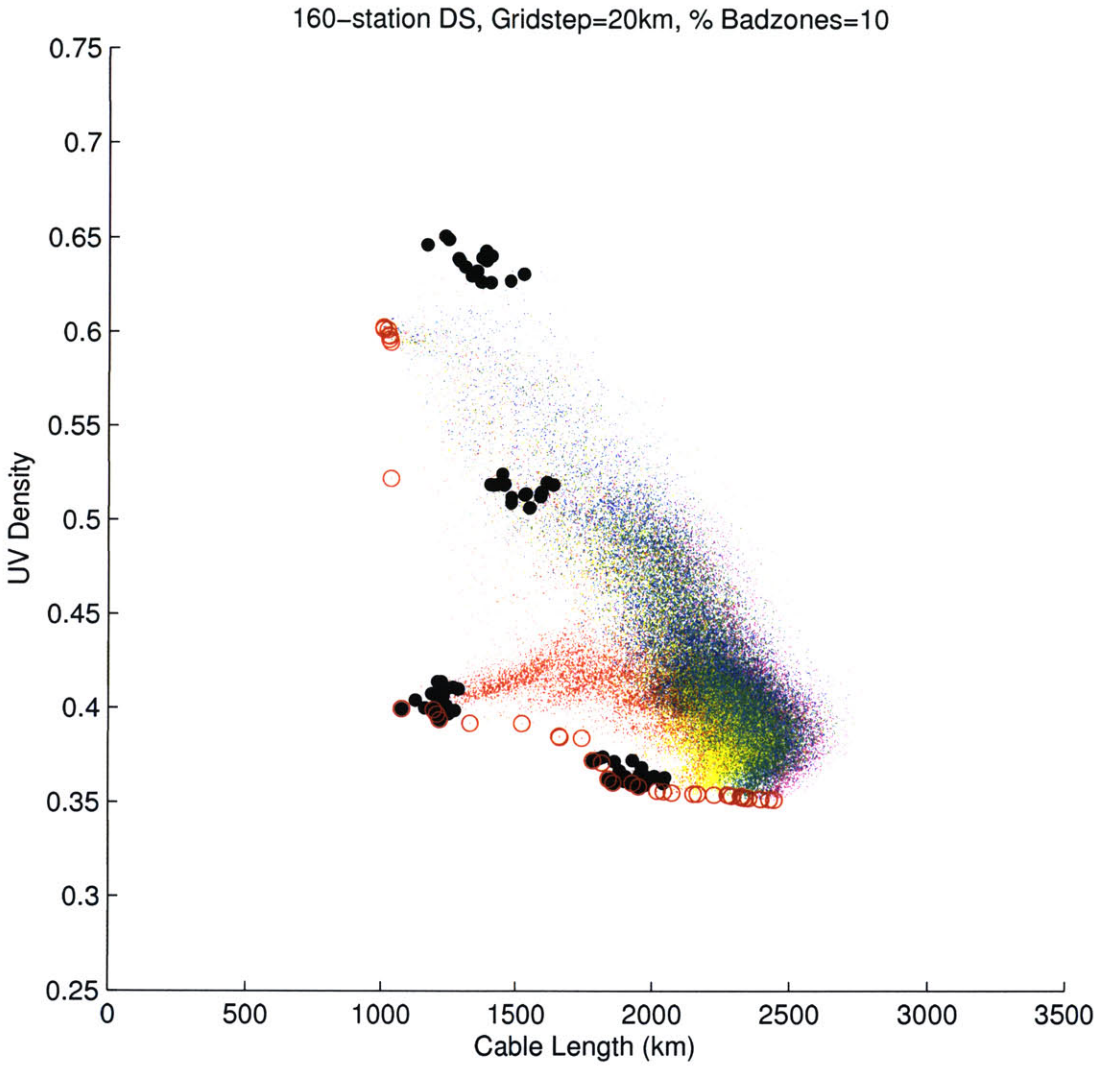


Figure A-72: 160-station, 20km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

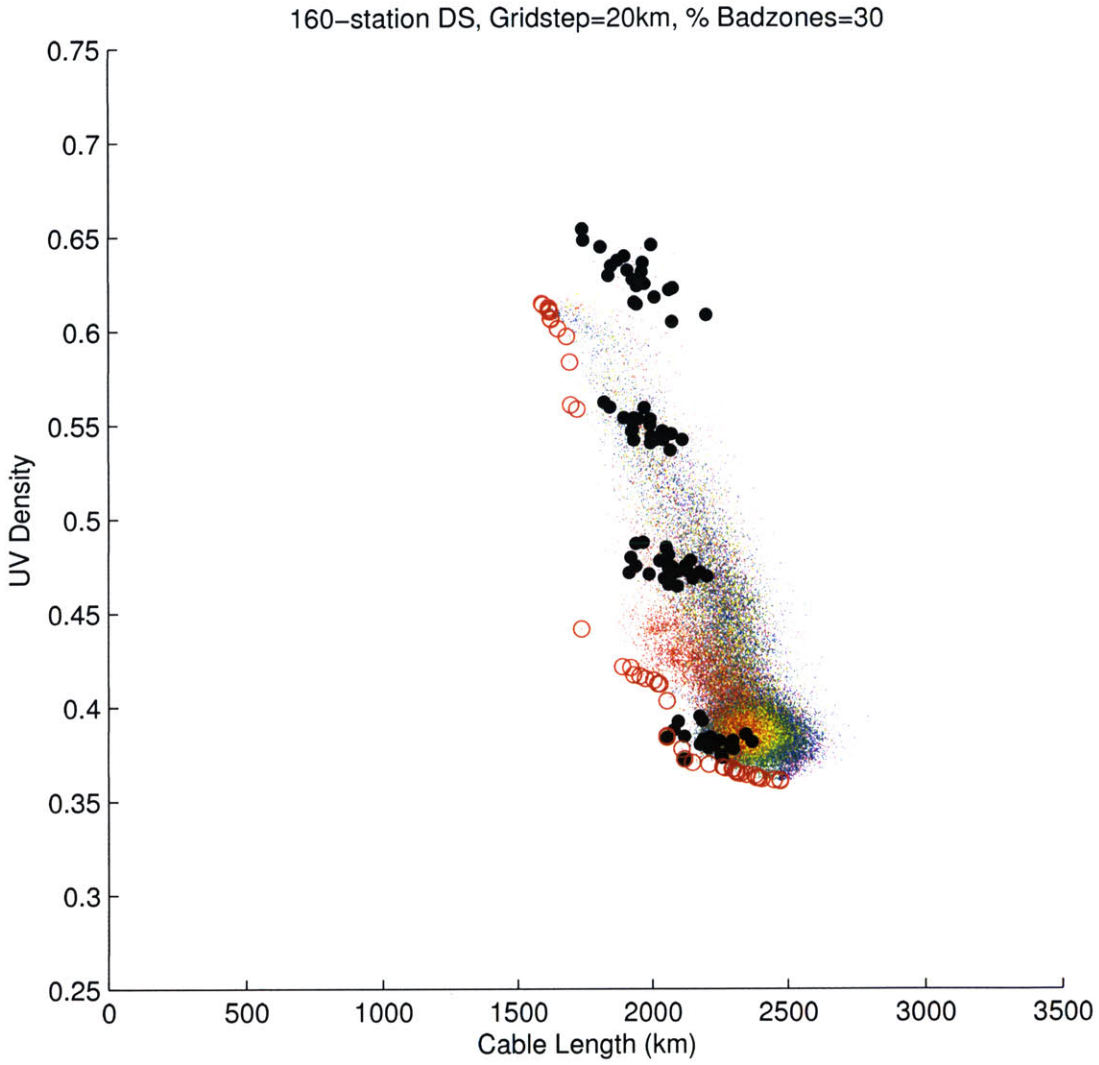


Figure A-73: 160-station, 20km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

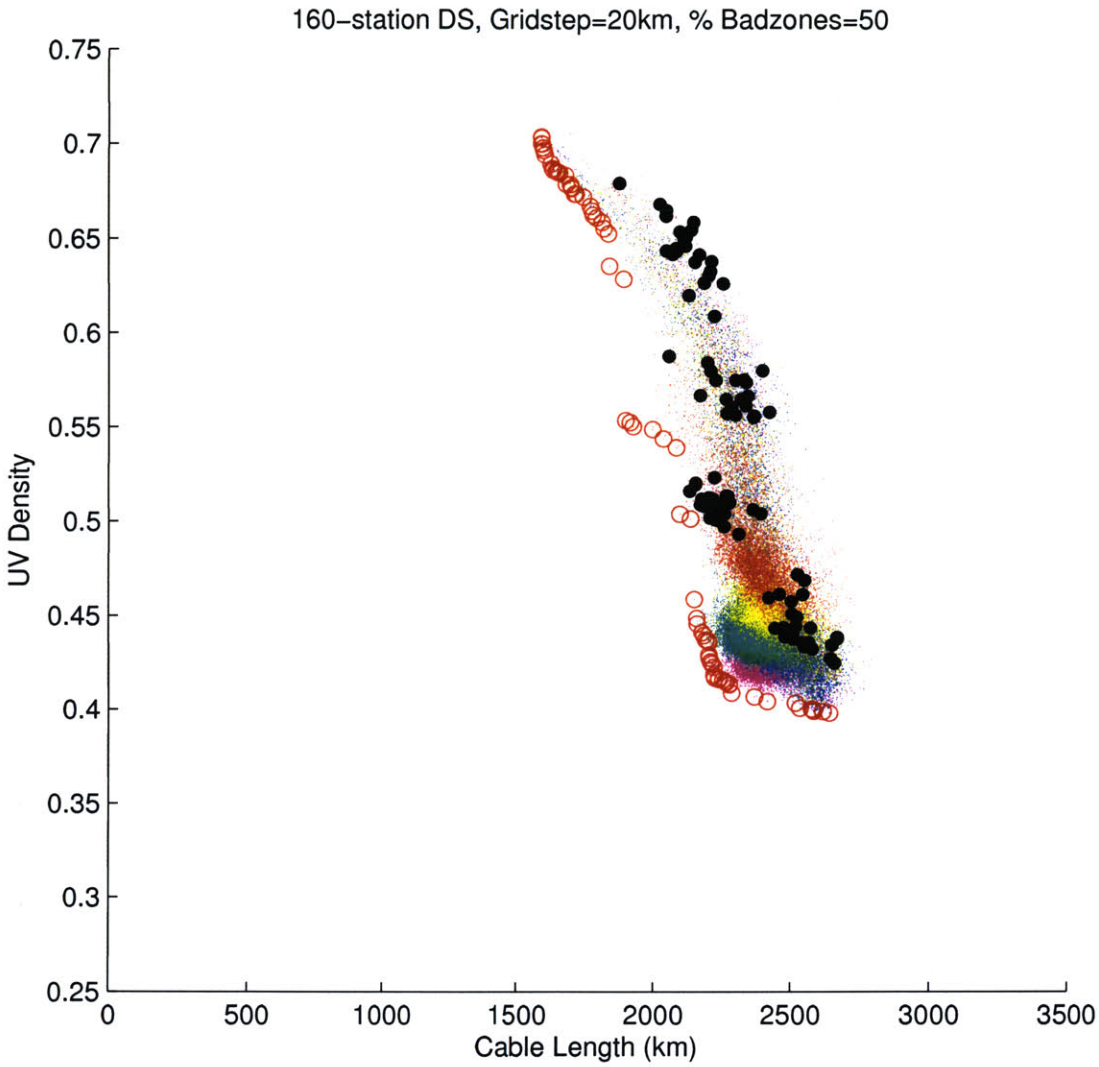


Figure A-74: 160-station, 20km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

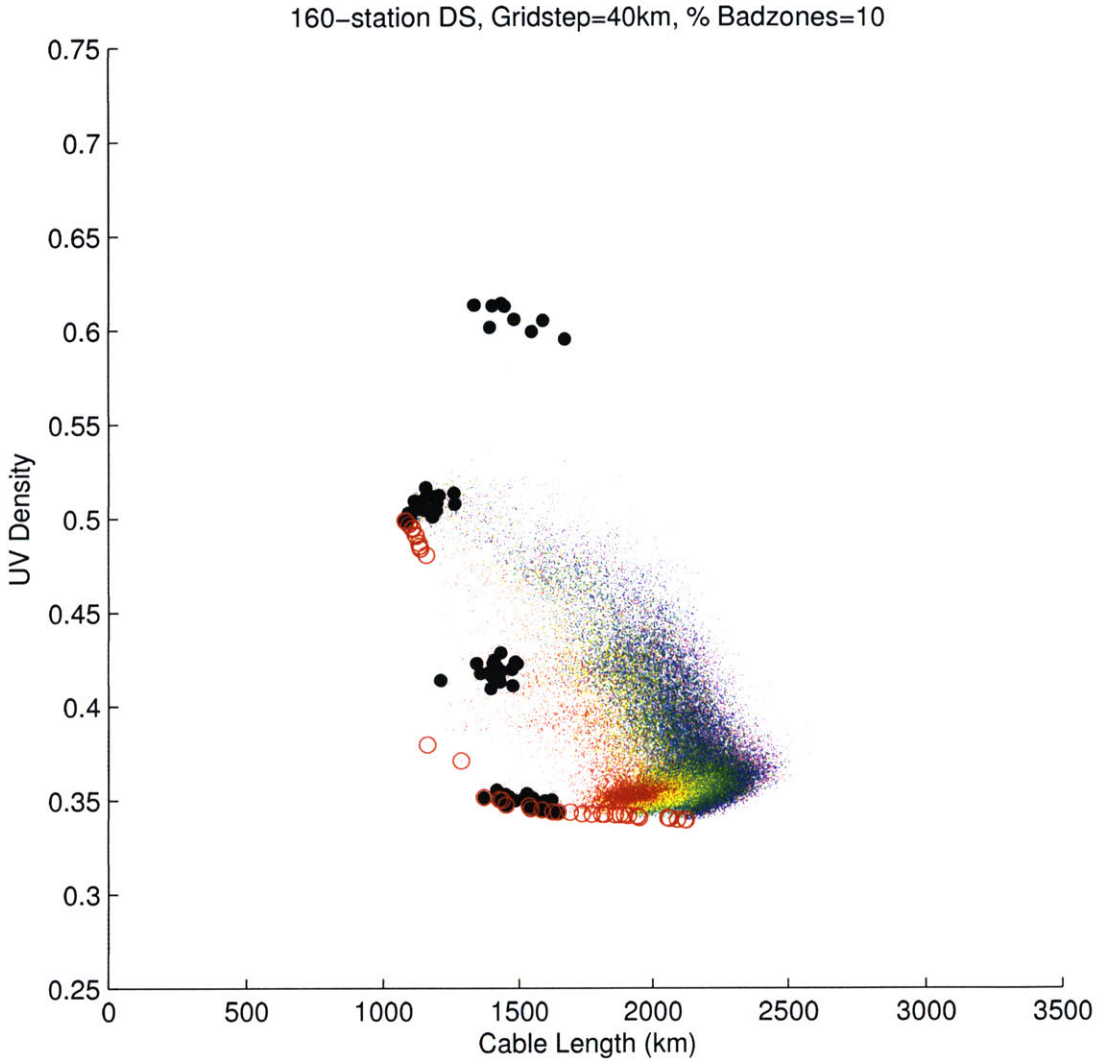


Figure A-75: 160-station, 40km *gridstep*, 10% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

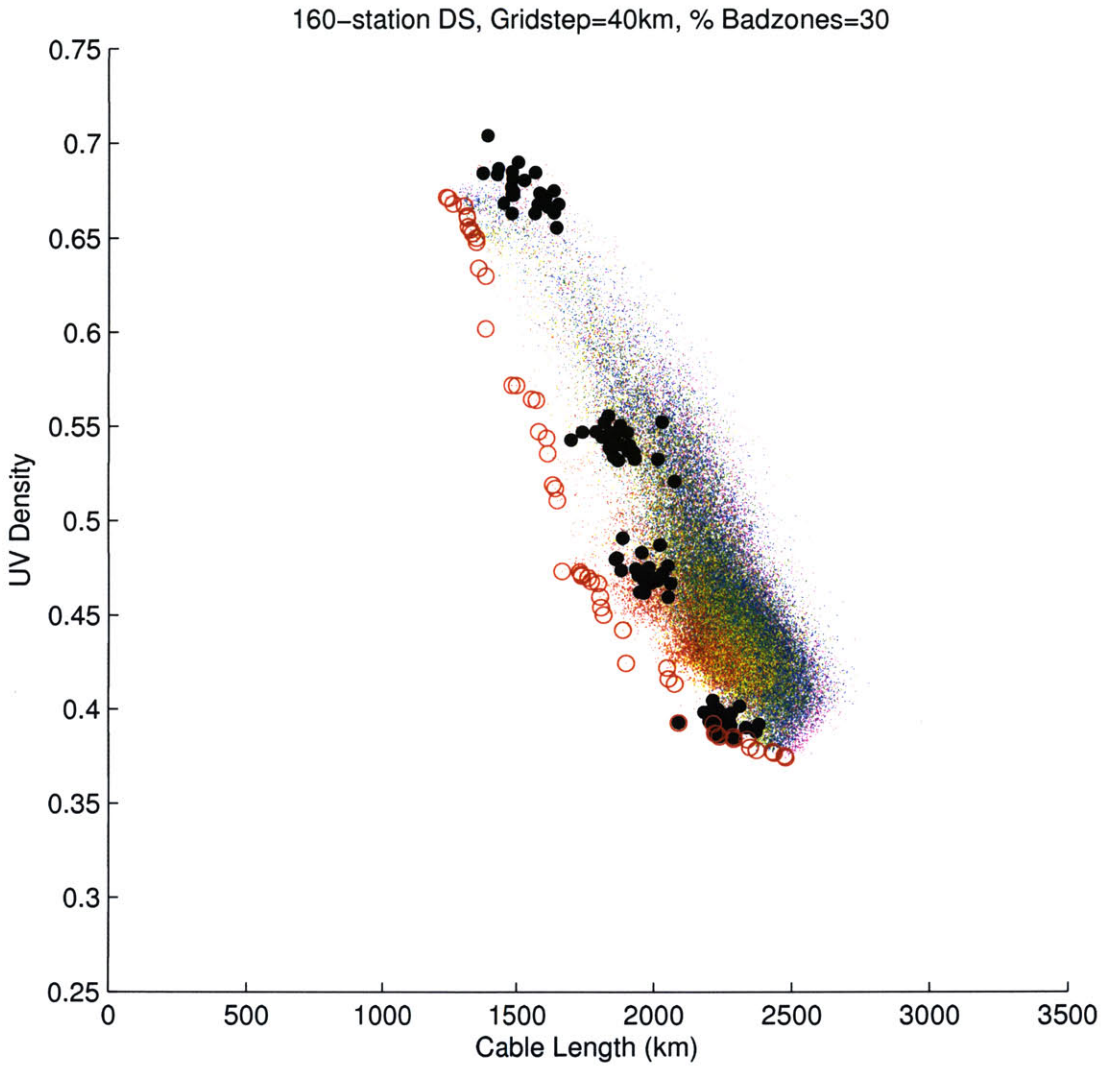


Figure A-76: 160-station, 40km *gridstep*, 30% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

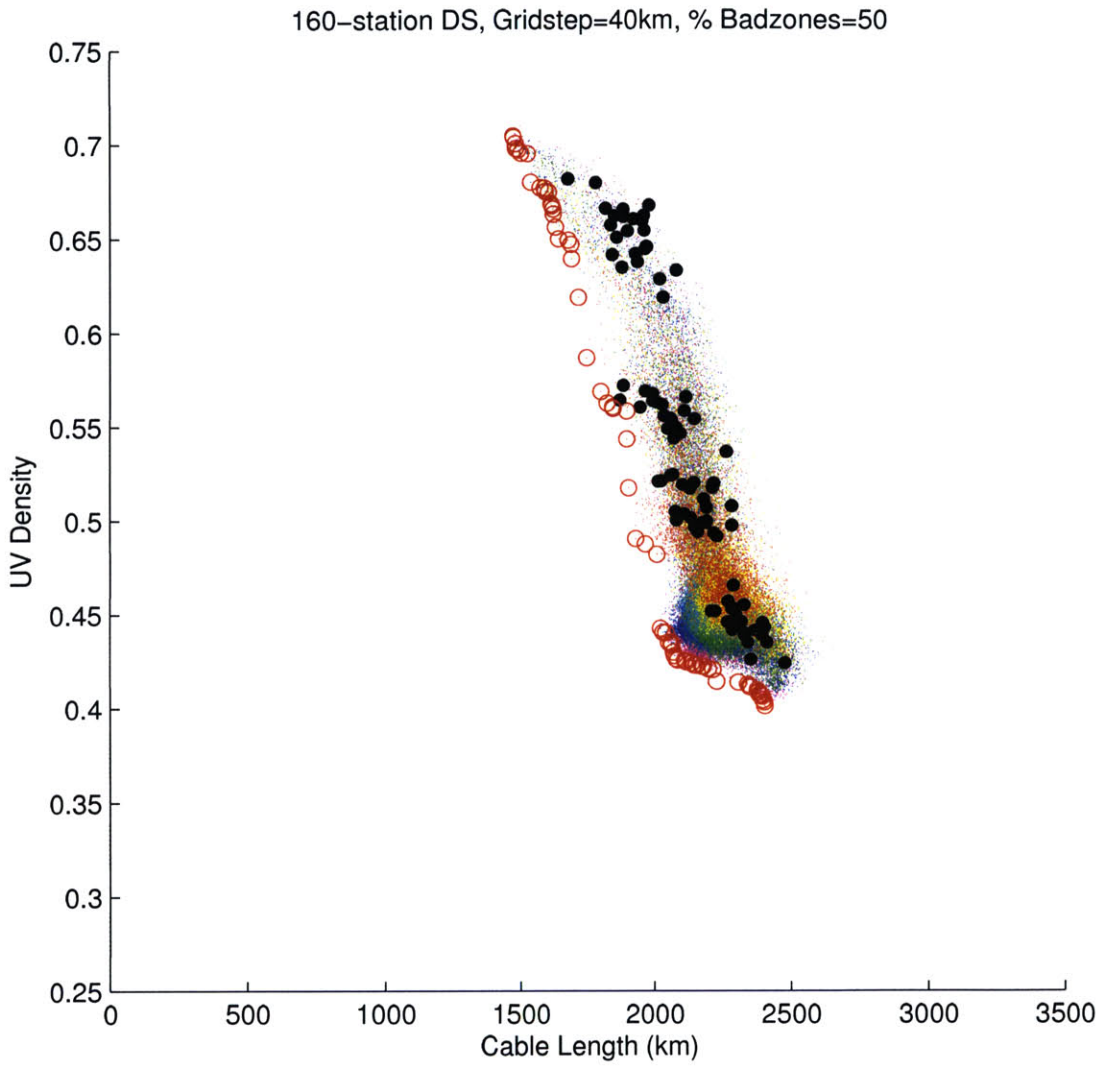


Figure A-77: 160-station, 40km *gridstep*, 50% badzones objective space. Black dots denote initial seeds of the population and are (from top-left to bottom-right) VLA-like configurations, triangles, Reuleaux triangles, and rings. Red dots correspond to the initial 10% of generations, yellows dots are the next 10% of generations, cyan the next 20%, green the next 20%, blue the next 20%, and magenta the final 20%. The non-dominated solutions are enclosed in red circles.

Appendix B

Code

Attached CD with GA_uvdens, GA_site, and all simulation runs used in this thesis.

Routines are also included for producing point spread function information.

All code was written using Matlab.