

The Life Cycle Energy-Water Usage Efficiency of Artificial Groundwater Recharge Via the Reuse of Treated Wastewater

A DISSERTATION PRESENTED
BY
ERIC DANIEL FOURNIER
TO
THE BREN SCHOOL OF ENVIRONMENTAL SCIENCE & MANAGEMENT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
ENVIRONMENTAL SCIENCE & MANAGEMENT

UNIVERSITY OF CALIFORNIA SANTA BARBARA
SANTA BARBARA, CALIFORNIA
JUNE 2015

APPROVED
BY
COMMITTEE CHAIR – DR. ARTURO A. KELLER

COMMITTEE MEMBER – DR. JAMES FREW

COMMITTEE MEMBER – DR. ROLAND GEYER

©2015 – ERIC DANIEL FOURNIER
ALL RIGHTS RESERVED.

The Life Cycle Energy-Water Usage Efficiency of Artificial Groundwater Recharge Via the Reuse of Treated Wastewater

ABSTRACT

This dissertation investigates the dynamic energy-water usage efficiencies of civil engineering projects involving the recharge of subsurface groundwater aquifers via the reuse of treated municipal wastewater. To this end, a three-component model has been developed. The first component uses a cartographic modeling technique known as Weighted Overlay Analysis (WOA) to develop a quantitative understanding of the location and extent of geographic areas that are suitable as sites for groundwater recharge in a given geographic context. The second component uses a Genetic Algorithm (GA) to address the multi-objective spatial optimization problem associated with locating corridors for the support infrastructure required to physically transport water from the treatment facility to the recharge site. The third and final component takes data about the anticipated recharge treatment source location, reuse destination location, and proposed infrastructure corridor location and uses them to populate a spatially explicit Life Cycle Inventory (LCI) model capturing all of the system's process energy consumption and material inputs. Five case studies involving the planning of new basin scale artificial recharge systems within the state of California are discussed.

Contents

o	INTRODUCTION	I
o.1	Research Objectives	2
o.2	The Energy-Water Nexus	3
o.3	Water Distribution Systems	6
o.4	Wastewater Treatment	8
o.5	Wastewater Recycling and Reuse	9
o.6	The Energy-Water Feedback Loop	10
i	SELECTING SUITABLE SITES	14
i.1	Selecting Suitable Sites for Artificial Groundwater Recharge Basins	15
i.2	Multi-Criteria Site Suitability Analyses	15
i.3	The MCSS Data Preprocessing Workflow	16
i.4	The MCSS Computational Workflow	17
i.5	The Software Toolset Repository Architecture	19
i.6	An Example Implementation	21
i.7	The Geographic Unit of Analysis	22
i.8	WOSS Model Outputs	23
2	LOCATING OPTIMAL CORRIDORS	25
2.1	Locating Optimal Corridors for Water Distribution Infrastructure	26
2.2	The Multi-Objective Corridor Location Problem	27
2.3	Genetic Algorithms	28
2.4	The MOGADOR Algorithm	29
2.5	The MOGADOR Data Structure	31
2.6	Initializing the MOGADOR Algorithm	32
2.7	An Example Pseudo-Random Walk	35
2.8	Initializing Problems with Large Decisions Spaces	37
2.9	Initializing Problems with Concave Decision Spaces	39
2.10	Measuring Initialization Performance	40
2.II	MOGADOR Algorithm Outputs	44

3	QUANTIFYING ENERGY-WATER RESOURCE USAGE	45
3.1	Quantifying Life-Cycle Energy-Water Resource Usage Efficiency	46
3.2	Life Cycle Assessment and Inventory Modeling	46
3.3	Wastewater Treatment Processes	47
3.4	Water Distribution Infrastructure	49
3.5	WWEST Recycled Water Reuse Life Cycle Inventory Model	51
3.6	WWEST Model Outputs	52
4	CASE STUDY IMPLEMENTATIONS	53
4.1	Santa Barbara Region	54
4.2	Oxnard Region	58
4.3	San Diego Region	63
4.4	Santa Ana – San Bernadino Region	68
4.5	Fresno – Tulare Region	73
4.6	Evaluating Algorithm Runtime Performance	83
4.7	Evaluating Algorithm Solution Quality	83
	REFERENCES	100

Listing of figures

1	Perspectives on the Energy-Water Nexus	4
2	Dimensions of the Energy-Water Nexus	5
3	Characteristics of Water Distribution Infrastructure	7
4	Energy Intensity of Water Distribution Infrastructure	8
5	Process Flow Diagram of Wastewater Treatment Methods	11
6	End Use Categories for Recycled Water	12
7	Water Intensity of Energy Production	13
1.1	Conceptual illustration of the input data preprocessing operations and workflow phases typical to most MCSS analyses	18
1.2	Control logic for MCSS input data preprocessing workflow.	19
1.3	Directory tree structure for the toolset repository. Filetypes required for input data and automatically generated as output data are shown.	20
1.4	Table of input data sources used in the case study MCSS model for artificial groundwater recharge applications.	22
1.5	Graphical Illustration of the WOSS Model Outputs for Reference HUC-10 Boundary	24
2.1	MOGADOR Algorithm Pseudocode	31
2.2	MOGADOR Algorithm Data Structure	32
2.3	Pseudo-Random Walk Algorithm Pseudocode	34
2.4	Pseudo-Random Walk Example	36
2.5	Conceptual Illustration of a Multi-Part Pseudo-Random Walk	38
2.6	Conceptual Illustration of a Convex Multi-Part Pseudo-Random Walk	40
2.7	Observing the Role of the Fixed Parameter q in Determining Individual Path Randomness	41
2.8	Observing the Role of Problem Size on Initialization Runtime	42
2.9	Observing the Characteristics of Multi-Part Pseudo-Random Walks	44
3.1	WWEST Model System Boundaries	52
4.1	Santa Barbara Region Overview (Filled in Black)	54
4.2	Santa Barbara Region Search Domain (Filled in Red)	55
4.3	Santa Barbara Region Destination Search Inputs: Slope Score (Blue:Low, Red:High)	55

4.4	Santa Barbara Region Destination Search Inputs: Geology Score (Blue:Low, Red:High)	56
4.5	Santa Barbara Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High)	56
4.6	Santa Barbara Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)	56
4.7	Santa Barbara Region Destination Search Outputs: Candidate Regions	57
4.8	Santa Barbara Region Proposed Corridor Endpoints	57
4.9	Santa Barbara Region Accessibility Based Objective Scores (Blue:Low, Red:High) .	58
4.10	Santa Barbara Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)	58
4.11	Santa Barbara Region Slope Based Objective Scores (Blue:Low, Red:High)	58
4.12	Santa Barbara Region Corridor Analysis Results	59
4.13	Santa Barbara Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview	59
4.14	Oxnard Region Overview (Filled in Black)	60
4.15	Oxnard Region Search Domain (Filled in Red)	61
4.16	Oxnard Region Destination Search Inputs: Slope Score (Blue:Low, Red:High) . .	61
4.17	Oxnard Region Destination Search Inputs: Geology Score (Blue:Low, Red:High) .	62
4.18	Oxnard Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High) .	62
4.19	Oxnard Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)	63
4.20	Oxnard Region Destination Search Outputs: Candidate Regions	63
4.21	Oxnard Region Proposed Corridor Endpoints	64
4.22	Oxnard Region Accessibility Based Objective Scores (Blue:Low, Red:High) . . .	64
4.23	Oxnard Region Land Use Based Disturbance Objective Scores (Blue:Low, Red:High)	65
4.24	Oxnard Region Slope Based Objective Scores (Blue:Low, Red:High)	65
4.25	Oxnard Region Corridor Analysis Results	66
4.26	Oxnard Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview . .	67
4.27	San Diego Region Overview (Filled in Black)	67
4.28	San Diego Region Search Domain (Filled in Red)	68
4.29	San Diego Region Proposed Corridor Endpoints	69
4.30	San Diego Region Accessibility Based Objective Scores (Blue:Low, Red:High) .	70
4.31	San Diego Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)	71
4.32	San Diego Region Slope Based Objective Scores (Blue:Low, Red:High)	72
4.33	San Diego Region Corridor Analysis Results	73
4.34	San Diego Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview .	74
4.35	Santa Ana – San Bernardino Region Overview (Filled in Black)	75
4.36	Santa Ana – San Bernardino Region Search Domain (Filled in Red)	75
4.37	Santa Ana – San Bernardino Region Destination Search Inputs: Slope Score (Blue:Low, Red:High)	76
4.38	Santa Ana – San Bernardino Region Destination Search Inputs: Geology Score (Blue:Low, Red:High)	76
4.39	Santa Ana – San Bernardino Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High)	77
4.40	Santa Ana – San Bernardino Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)	77

4.41	Santa Ana – San Bernadino Region Destination Search Outputs: Candidate Regions	78
4.42	Santa Ana – San Bernadino Region Proposed Corridor Endpoints	78
4.43	Santa Ana – San Bernadino Region Accessibility Based Objective Scores (Blue:Low, Red:High)	79
4.44	Santa Ana – San Bernadino Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)	79
4.45	Santa Ana – San Bernadino Region Slope Based Objective Scores (Blue:Low, Red:High)	80
4.46	Santa Ana – San Bernadino Region Corridor Analysis Results	81
4.47	Santa Ana – San Bernadino Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview	82
4.48	Fresno – Tulare Region Overview (Filled in Black)	82
4.49	Fresno – Tulare Region Search Domain (Filled in Red)	83
4.50	Fresno – Tulare Region Destination Search Inputs: Slope Score (Blue:Low, Red:High)	84
4.51	Fresno – Tulare Region Destination Search Inputs: Geology Score (Blue:Low, Red:High)	85
4.52	Fresno – Tulare Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High)	86
4.53	Fresno – Tulare Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)	87
4.54	Fresno – Tulare Region Destination Search Outputs: Candidate Regions	88
4.55	Fresno – Tulare Region Proposed Corridor Endpoints	89
4.56	Fresno – Tulare Region Accessibility Based Objective Scores (Blue:Low, Red:High)	90
4.57	Fresno – Tulare Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)	91
4.58	Fresno – Tulare Region Slope Based Objective Scores (Blue:Low, Red:High)	92
4.59	Fresno Region Corridor Analysis Results	93
4.60	Fresno Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview	94
4.61	Algorithm Runtime Performance for Each of the Five Case Study Regions for Three Population Sizes	95
4.62	Algorithm Convergence Rates for Each of the Five Case Study Regions for Three Population Sizes	96
4.63	Comparison of the Along Path Distance and Cumulative Objective Scores between the Solution Corridors and the Euclidean Shortest Corridors	97

THIS DISSERTATION IS DEDICATED TO THE MEMORY OF JULIO ARNOLD MUNIZ.
HE WAS MY BEST FRIEND.

Acknowledgments

I WOULD LIKE TO THANK my father Paul, my mother Patricia, my brother Bryan, and my girlfriend Tessa for their continued support of my education. I would also like to thank my PhD Committee members, Dr. Arturo Keller, Dr. James Frew, and Dr. Roland Geyer for their respective contributions to the development and successful completion of this project. Finally, I would like to thank Dr. Charles Dana Tomlin for inspiring me to pursue a doctorate in this field of study through his work, teaching, eloquence, and integrity.

*We have arranged a civilization in which most crucial
elements profoundly depend upon science and technology.*

Carl Sagan (1934-1996)

0

Introduction

0.1 RESEARCH OBJECTIVES

THE OVERARCHING RESEARCH OBJECTIVE of this dissertation can be stated as follows: develop an integrated modeling framework, incorporating data depicting local geographic context, that is capable of quantifying the life-cycle energy-water usage efficiency of proposed new infrastructural systems supporting the reuse of treated municipal wastewater for the purpose of artificial groundwater recharge. In order to achieve this goal the following three canonical problems were identified and addressed in sequential order:

1. Given multiple independent objectives – propose a systematic method for selecting sites that are suitable for the development of new artificial groundwater recharge infrastructure.
2. Given multiple independent objectives – propose a systematic method for routing optimal corridors for new water distribution infrastructure linking a designed treated effluent source to a designated recharge destination.
3. Given a fixed influent flow rate, a set of mean influent pollutant concentrations, a set of maximum effluent pollutant concentrations, and the topological structure of the effluent distribution network – propose a systematic method for determining the life-cycle energy-water usage efficiency of an integrated wastewater reuse system

The remainder of Chapter (0) provides an in-depth background discussion of the both the academic and social relevance of the stated research objective. Subsequent Chapters (1-3) are organized with respect to the various independent research activities that were undertaken to address each of the three canonical problems listed above. The final Chapter (4) provides an in depth analysis of a set five case study implementations that were developed to assess the framework's efficacy in satisfying the stated research objective.

0.2 THE ENERGY-WATER NEXUS

NEARLY ALL MODERN INDUSTRIALIZED SOCIETIES rely upon energy generation technologies which are derivative of a thermodynamic process known as a heat engine. In a typical heat engine the chemical energy stored within a fuel source such as coal, petroleum, or natural gas, must be first be released as ambient thermal energy through the process of combustion. The heat engine then, by virtue of its design, converts this ambient thermal energy into mechanical energy for the purpose of performing some sort of meaningful work – i.e. generating electricity.

The history of the advancement of the human species is a story which can be cast in terms of the progressive discovery new, higher density chemical energy stores and their enhanced exploitation via the development of new, ever more sophisticated heat engines. Interestingly however, is the fact that nowhere in our history has there ever occurred a single substantial deviation in the choice of the working fluid that actually performs the crucial energy conversion process within a heat engine: water. For all of the advances which have been made in terms of improved fuel processing, boiler and combustion chamber design, etc., water has and likely will continue to remain stubbornly positioned as a critical component of nearly all major commercial scale energy systems; and, in particular, those involved with the production of electricity.

Another technological system which can also be viewed as a foundation pillar of modern industrialized societies is the mechanized disposal of human and animal wastes via engineered sewage conveyance and treatment processes. Here again, the advancement of the human species might alternatively be cast in terms of the progressive improvement in the reliability and efficiency of these systems over time. And, what is more, just as the advancement of our energy system appears to be bounded by the physical properties of the water, so too has the advancement of wastewater management been similarly constrained.

For example, in terms of water treatment processes, the upper bound on process efficiency is determined by the Second Law of Thermodynamics. This law states that, for any closed system, there is a tendency for the entropy of that system to increase over time. *Ceteris paribus* wastewater, a heterogeneous mixture, exists in a higher entropy state than pure water does. This means that any attempt to purify a polluted wastewater stream must necessarily incur the cost of some energy input to facilitate the requisite reduction in entropy.

In terms of water distribution processes, the key physical determinant of energy efficiency stems from the density of water. At 1000 kg/m^3 , considerable energy must be expended whenever large quantities of water must be moved over a distance or lifted against an elevation gradient. What is more, many treatment processes utilize pressurized sieving techniques where water is physically pushed through a porous membrane in order to mechanically filter out dissolved pollutant species. As a consequence of this practice, the energy inputs required to overcome the previously mentioned entropic gradient are often supplied for the immediate purpose of moving quantities of water from place to place.

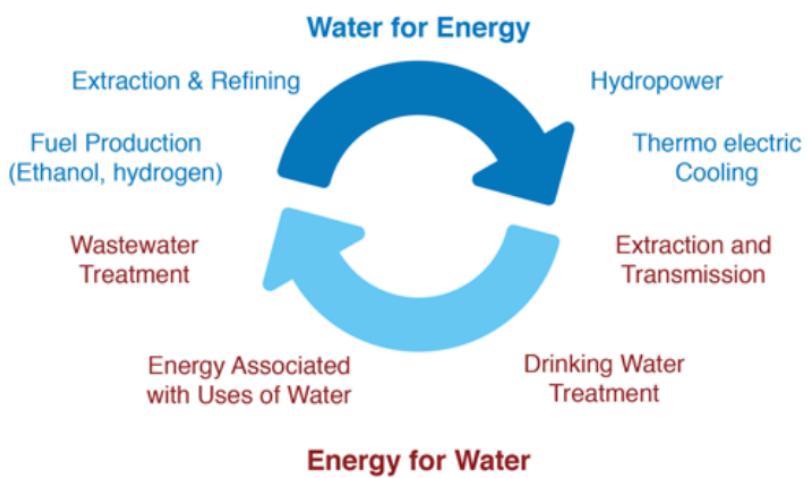


Figure 1: Perspectives on the Energy-Water Nexus

The energy-water nexus is a term which has emerged from within the academic research community to describe these types of dynamic interrelationships which are inherent to our energy and water systems. There are two perspectives from which the energy-water nexus can be alternatively studied. The first emphasizes the *water for energy* dimension, and is generally concerned with the study of processes and technologies that are involved with the direct withdrawal and consumption of water for the production of both primary and final energy resources. The second of these perspectives, focuses alternatively on the *energy for water* dimension; investigating processes and technologies which consume energy for the purpose of transmitting or purifying freshwater resources.

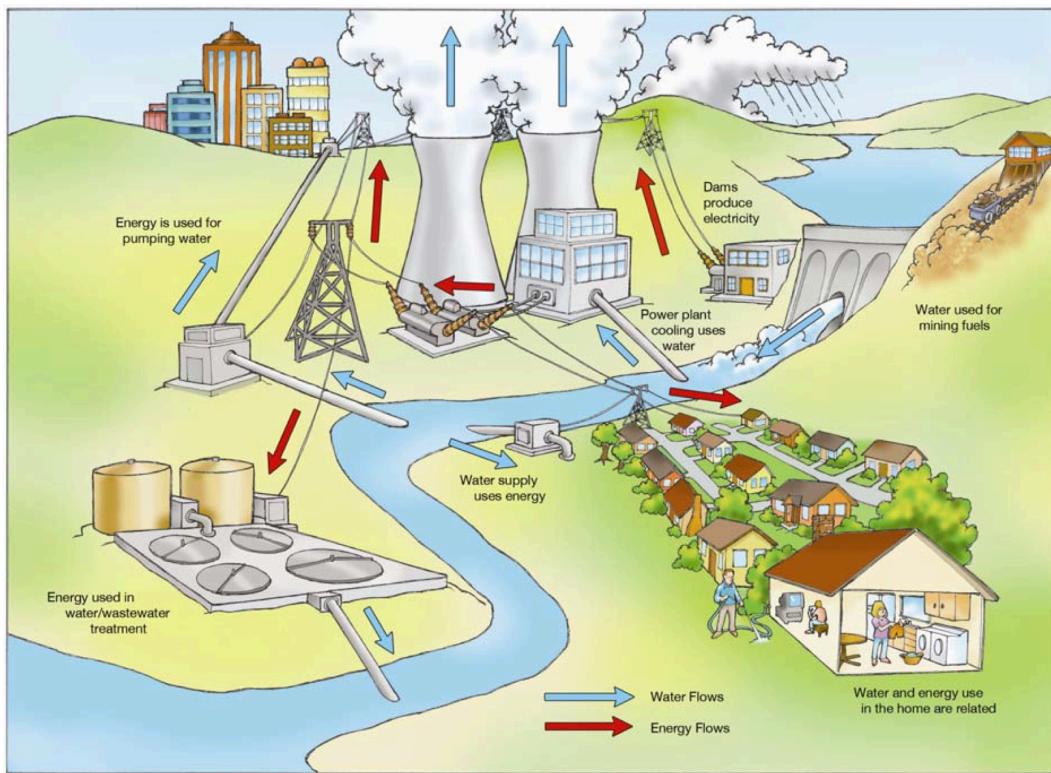


Figure 2: The dimensions of the energy-water nexus

Here in the United States, 50% of total annual freshwater withdrawals are used for the cooling of thermoelectric power plants. Alternatively, 4% of the nation's total energy consumption is dedicated

to the transmission and purification of water and wastewater. While these national figures speak to the overall significance of this issue, the situation becomes more acute when one begins to consider different regional contexts.

The criticality of the energy-water nexus becomes greatly exacerbated in those areas where either water is scarce, energy is scarce, or both. Unfortunately, the state of California suffers from both of these conditions to varying degrees, making the energy-water nexus a frequent source of regional interest within both the academic research and socio-political circles. For example, in a 2005 report published by the California Energy Commission (CEC) it was found that 19% of the electricity and 32% of the natural gas consumed within the entire state were used for purpose directly related to the supply and treatment of freshwater resources.

0.3 WATER DISTRIBUTION SYSTEMS

One of the main drivers for this tremendous energy consumption within the state of California is the large scale transfer of freshwater resources between physically distinct hydrologic basins. California is crossed latitudinally by a massive network of interconnected hydraulic engineering projects including pipelines, aqueducts, reservoirs, and pump stations. These systems, which have been funded by a mixture of Federal, State, and Local agencies, were designed to reconcile discontinuities between the spatial and temporal distributions of the supply and demand for freshwater resources within the state.

Inter-basin transfers typically involve the movement of water against a considerable elevation gradient. Due to water's previously mentioned high density, there are substantial energetic costs associated with operating the infrastructure required to facilitate these transfers. For example, the bar graph to the right of Figure 4 compares the energy intensity of several different sources of municipal water within the state of California. According to this research water resources which are supplied



Figure 3: The geographic extent of water distribution infrastructure in the state of California

via inter-basin transfer, either through branches of the State Water Project or through the Colorado River Aqueduct, rank very poorly in terms of energy usage efficiency relative to a number of other water supply systems.

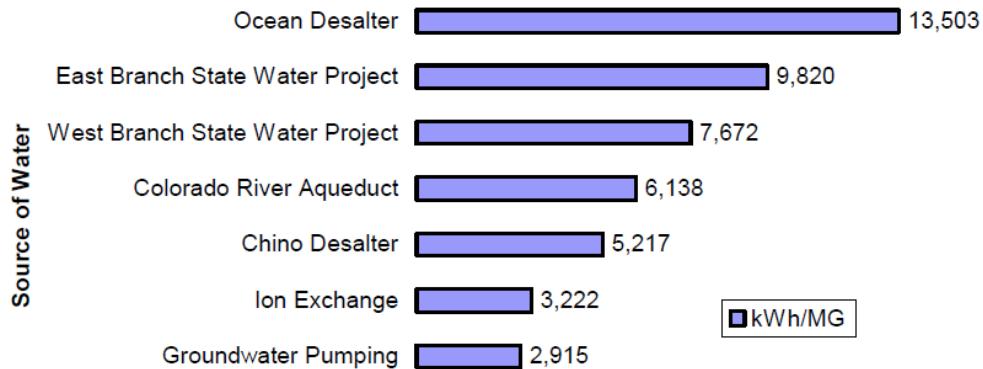


Figure 4: The geographic extent of water distribution infrastructure in the state California

0.4 WASTEWATER TREATMENT

The term *wastewater treatment* refers to a set of physical and chemical processes that have been individually designed and are collectively composed for the purpose of removing a suite of physical, chemical, and/or biological contaminants from a quantity of water. The operational goal of a wastewater treatment process is typically defined in terms of a set of desired effluent pollutant concentrations that are sufficiently low to enable that effluent to be used for a specific end-use application. Crucially implicit in this definition therefore, is the notion that the precise combination and configuration of the individual, atomic treatment processes, will vary on the basis of the purity requirements associated with the anticipated end-use application. Figure 5 provides a high level view of the separate treatment processes that are typical to modern wastewater treatment facilities; assigning them to a somewhat informal hierarchy consisting of: primary treatment, secondary treatment, and tertiary treatment.

The collection of processes described in Figure 5 represent the required suite of treatment operations necessary to produce effluent water that is suitable for groundwater recharge applications

(indirect potable reuse) given a normal stream of influent municipal wastewater. Typically, wastewater treatment plants (WWTPs) – ones that are producing treated effluent for release into natural waters – are only required to provide up to a secondary treatment level; which, on this diagram, corresponds to the fourth row of processes from the top. As the figure shows, providing treated effluent for reuse applications necessitates the implementation of as many as seven additional layers of tertiary treatment in order to manage issues such as suspended colloids, disinfection, and phosphorous removal.

This hierarchy of treatment processes depicted in 5 is only meant to be representative of the types of processes that are typically required for reuse applications. In reality however, neither the rigid segregation of treatment processes nor the association of each treatment level with a set of designated end-uses have been precisely codified into a coherent regulatory framework at the Federal level. In general, primary and secondary treatment are Federally mandated as part of the generic WWTP under provisions of the Clean Water Act. However, tertiary treatment processes and requirements for their application in the context of different desired end-use applications, are, at present, regulated at the state or local levels in a more ad-hoc manner. Figure 6 attempts to illustrate the complexity of this landscape by mapping treatment levels to different end-use types, with annotations where additional restrictions may apply.

0.5 WASTEWATER RECYCLING AND REUSE

Over the past ten years the reuse of treated wastewater has become the fastest growing source of new water supply for municipal water districts in the Western United States. There are a variety of reasons for this growth. Some districts enjoy the degree of self determinism that a reuse system affords them; particularly in cases where they are beholden to the whims of a third party water wholesale provider or are ham strung by more senior water rights holders elsewhere within their

basin. In general however, the primary driving force behind this growth trend has been the fact that treated municipal wastewater is perceived to be an energy efficient water supply alternative for a number of low quality uses.

If one looks at current reuse systems, this assumption appears to be safe. This is because the vast majority of these systems are structured such that the destination location for the treated effluent is situated in close proximity to the WWTP. An illustrative example of such a situation might be the use of wastewater which had been subjected to basic secondary treatment for the irrigation of a nearby cemetery or golf course. These types of reuse systems can be thought of as "low hanging fruit" because the only additional energy process based energy expenditures associated with their operation come from whatever tertiary treatment processes must be added into the WWTP operations to meet the effluent purity requirements associated with the designated end-use.

More recently however, many municipalities have begun to actively investigate the possibility of expanding the both the scale and extent of their reuse operations: capturing larger fractions of their WWTP plants' throughput and distributing the treated effluent to a more diverse portfolio of end-use recipients.

o.6 THE ENERGY-WATER FEEDBACK LOOP

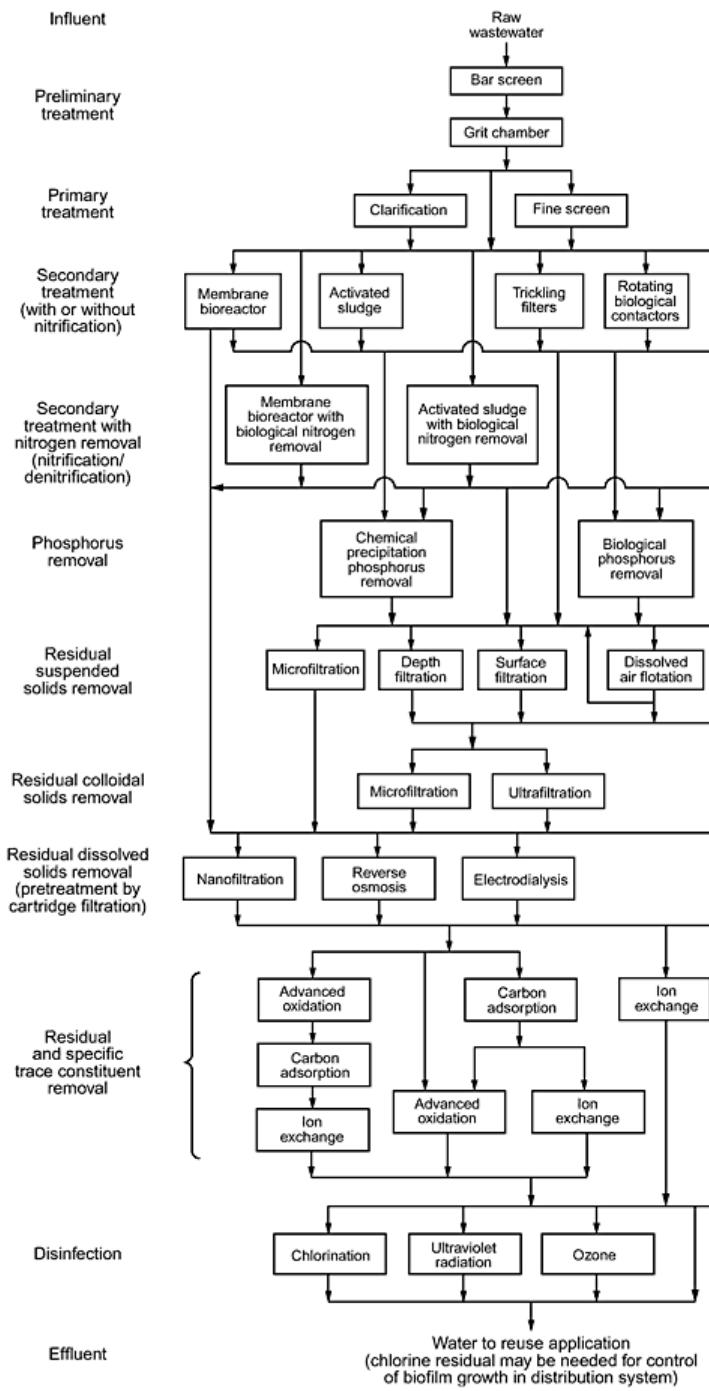


Figure 5: Process flow diagram of various wastewater treatment methods

Types of Use	Treatment Level		
	Disinfected Tertiary	Disinfected Secondary	Undisinfected Secondary
Urban Uses and Landscape Irrigation			
Fire protection	<input checked="" type="checkbox"/>		
Toilet & urinal flushing	<input checked="" type="checkbox"/>		
Irrigation of parks, schoolyards, residential landscaping	<input checked="" type="checkbox"/>		
Irrigation of cemeteries, highway landscaping		<input checked="" type="checkbox"/>	
Irrigation of nurseries		<input checked="" type="checkbox"/>	
Landscape impoundment	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> *
Agricultural Irrigation			
Pasture for milk animals		<input checked="" type="checkbox"/>	
Fodder and fiber crops		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Orchards (no contact between fruit and recycled water)		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Vineyards (no contact between fruit and recycled water)		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Non-food bearing trees			<input checked="" type="checkbox"/>
Food crops eaten after processing		<input checked="" type="checkbox"/>	
Food crops eaten raw	<input checked="" type="checkbox"/>		
Commercial/Industrial			
Cooling & air conditioning - w/cooling towers	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> *
Structural fire fighting	<input checked="" type="checkbox"/>		
Commercial car washes	<input checked="" type="checkbox"/>		
Commercial laundries	<input checked="" type="checkbox"/>		
Artificial snow making	<input checked="" type="checkbox"/>		
Soil compaction, concrete mixing		<input checked="" type="checkbox"/>	
Environmental and Other Uses			
Recreational ponds with body contact (swimming)	<input checked="" type="checkbox"/>		
Wildlife habitat/wetland		<input checked="" type="checkbox"/>	
Aquaculture	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> *
Groundwater Recharge			
Seawater intrusion barrier	<input checked="" type="checkbox"/> *		
Replenishment of potable aquifers	<input checked="" type="checkbox"/> *		

*Restrictions may apply

Figure 6: End Use Categories for Recycled Water

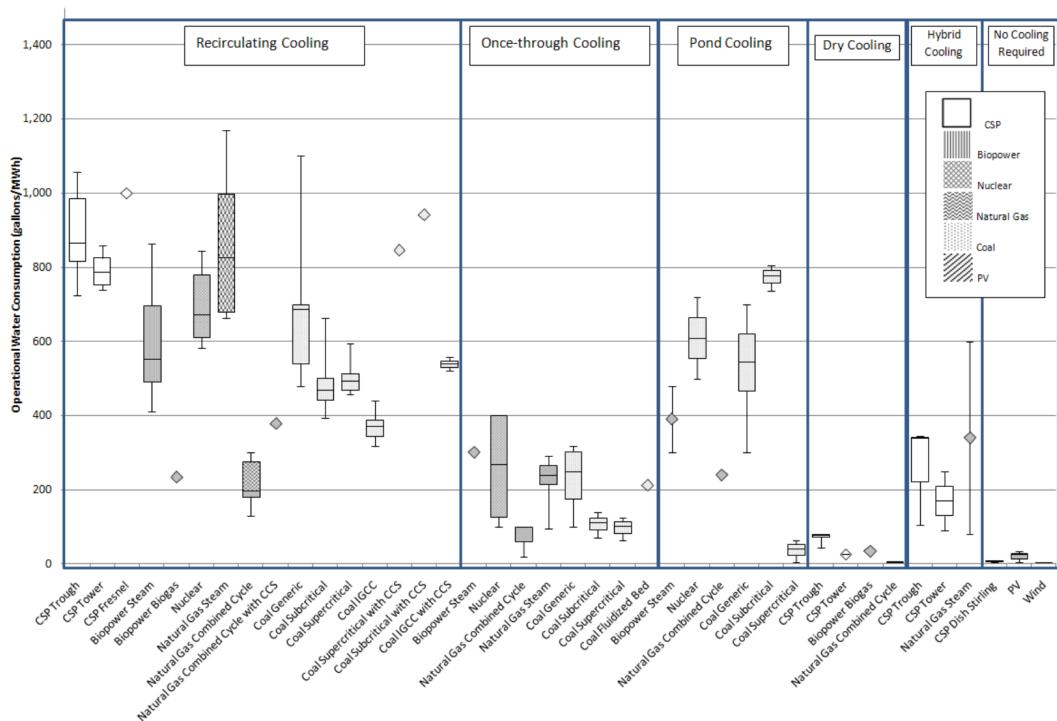


Figure 7: The water consumption intensity, measured in terms of water consumed per unit power produced, for various electricity generation technologies

Do not lose your faith. A mighty fortress is our mathematics. It shall rise to the occasion. It always has.

Stanislaw Ulam (1909-1984)

1

Selecting Suitable Sites

1.1 SELECTING SUITABLE SITES FOR ARTIFICIAL GROUNDWATER RECHARGE BASINS

THE FIRST PRINCIPLE CHALLENGE which must be overcome in an attempt to quantify the life-cycle energy water usage efficiency of a hypothetical water reuse system involving artificial groundwater recharge is the need to develop of a systematic method for selecting sites that are suitable for the construction of the requisite recharge basins. This problem of *selecting suitable sites*, relative to one or more criteria of suitability, is one which is extremely common within the domain of Geographic Information Science (GIScience). And indeed, the need to develop consistent methodologies and capable tools for achieving this purpose were among the core research goals which initially stimulated the early development of the field.

1.2 MULTI-CRITERIA SITE SUITABILITY ANALYSES

One methodology which has emerged as a reliable means of approaching this type problem, and the one which was adopted for the purposes of this dissertation, is a procedure known as multi-criteria site suitability (MCSS) analysis. MCSS analyses mathematically combine two or more input geographic data layers that each correspond to some independent measure of site suitability for a given landuse application³. The output of this MCSS computation is a single geographic data layer in which the value at each location represents a composite measure of overall site suitability relative to all of the independent criteria, simultaneously^{16,9}.

MCSS analyses are typically conducted using geographic information that has been stored in a continuous raster format. This means that prior to conducting this type of analysis each of the input geographic data layers that are to be used must be preprocessed relative to some reference raster format so as to ensure the feasibility and consistency of the MCSS computation. For example, in order for the computation to be feasible: all of the input data layers must have the same number of cells

and occupy that same geographic extent. Similarly, in order for the computation to be consistent: the ordinality and the scaling of the values in each raster must accurately reflect the relative weighting and directionality of each independent suitability criterium.

Geographic information system (GIS) software packages are commonly used to conduct both these types of data preprocessing operations as well as the MCSS computation itself^{20,21}. This is because they provide pre-built functions which facilitate the import of spatial data layers from disparate sources as well as the manipulation of geographic data layers such that they satisfy the previously mentioned feasibility and consistency constraints. Often, the MCSS analyses itself is not the endpoint goal of a given research effort however. Many times, MCSS analyses are used as inputs to some other, more complex, numerical optimization model⁵. This situation is frequently encountered in the fields of operations research (OR) and location science (LS) where MCSS outputs are used to derive network topologies or linear programming constraints for optimization problems related to vehicle routing, flow maximization, or facility location^{17,6}.

1.3 THE MCSS DATA PREPROCESSING WORKFLOW

The generic data preprocessing workflow for MCSS analysis involves one or more of the three phases illustrated conceptually in Figure 1.1. First, all of the input spatial data layers must be checked to determine whether or not they possess the same coordinate system. If they do not, a single reference coordinate system must be chosen by the analyst to function as the standard for all of the input layers. The choice of this reference coordinate system is generally driven by the scale of spatial domain involved as well as the desired tradeoff between distance versus areal measurement error. The output of this *Reproject* operation is a set of duplicate data layers that are all projected as the reference coordinate system.

The second phase of the workflow involves clipping the various input data layers on the basis

of their overlap with some reference spatial extent. This reference extent may correspond to the boundaries of a single input layer or may be arbitrarily designated by the analyst. The output of this *Clip* operation is a set of duplicate data layers that all have the same spatial extent as the reference extent.

The third and final phase of the workflow involves rasterizing all of the input layers such that they have the same cell size and cell alignment. During this phase, input data layers that are stored using a different geographic data model must be algorithmically converted into a raster based representation. As part of this algorithmic conversion process, a reference cell size, usually corresponding to the largest cell size contained within the input data layers, is used as a reference. The output of this *Rasterize* operation is a set of duplicate layers that all share the same cell size and cell alignment as the designated reference.

1.4 THE MCSS COMPUTATIONAL WORKFLOW

Instead of executing the MCSS data preprocessing steps manually for each of the five case study regions that were going to be investigated as part of this dissertation, the decision was made to implement a more generic data preprocessing software framework which enable any researcher – not necessarily the author of this dissertation – to conduct a similar analysis for any region in which sufficient input data was available. These tools, which shall be described in subsequent sections, were implemented in the the MATLAB® programming environment and have been made publicly available as a source code repository hosted at: <https://github.com/ericdfournier/WOSS>. The control flow logic which guided their development is described in the pseudocode in Figure 1.2.

Prior to the initiation of any MCSS analysis the researcher must provide a set of raw spatial data input files and designate a set of spatial reference criteria. Once these requirements are met, all of the input spatial data files are then subjected to a sequence of conditional statements. Depending

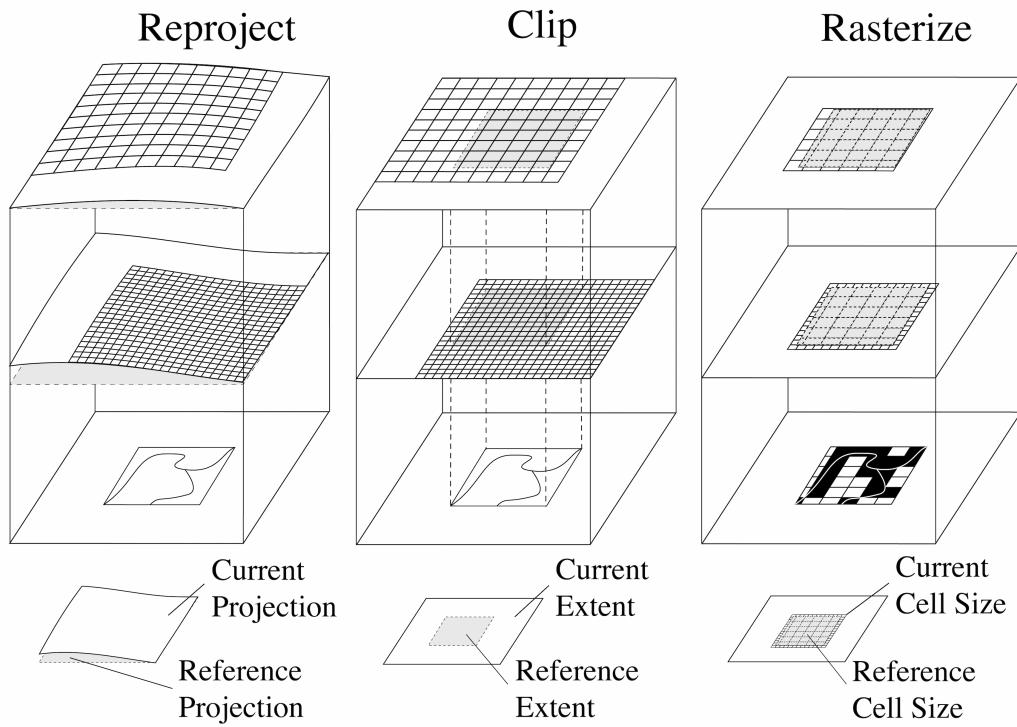


Figure 1.1: Conceptual illustration of the input data preprocessing operations and workflow phases typical to most MCSS analyses

upon the result of these conditional statement evaluations various transformation functions are then sequentially applied to the input data files so as to produce a set of outputs whose projection, spatial extent, cell size, and cell alignment all match a set of designated spatial reference criteria.

The current version of the toolset only supports the reprojection of input spatial data layers that are represented using geographic coordinates – i.e. data stored in latitude & longitude coordinate space). This constraint not only limits the directionality of the reprojection operation but also greatly simplifies the spatial interpolation routines required for the rasterization process. The authors plan to lift this restriction in future versions of the toolset as the MATLAB® language’s native support for forward map projection as well as the automated parsing of standard formatted spatial reference data strings improves.

```

1: procedure PREPROCESS INPUT DATA
2:   for all input do
3:     if inputprj ≠ referenceprj then
4:       input = Reproject(input)
5:     else if inputext ≠ referenceext then
6:       input = Clip(input)
7:     else if isRaster(inputtype) ≠ True then
8:       input = Rasterize(input)
9:     else if inputcs ≠ referencecs then
10:      input = Resize(input)
11:    end if
12:    output = input
13:  end for
14:  return output
15: end procedure

```

Figure 1.2: Control logic for MCSS input data preprocessing workflow.

Following the preprocessing of the spatial data inputs, the next phase of the MCSS modeling process is the user guided reclassification of the data values in each layer into a quantitative measure of suitability for the land use application in question. Routines are provided in the toolset to facilitate this process. Among these are an automated histogram equalization based reclassification procedure which assigns suitability values ensuring an even distribution of all the values contained within some range across all of the areas within the spatial data layer. Other tools allow the user to manually specify the range of the bins used for the reclassification of raw input data values to site suitability rankings.

1.5 THE SOFTWARE TOOLSET REPOSITORY ARCHITECTURE

The WOSS toolset is publicly available via the GitHub repository hosted at: <https://github.com/ericdfournier/WOSS>. The directory structure of this repository is illustrated in Figure 1.3. Below the top level root directory are four standalone files: (1) a *LICENSE.md* file, (2) a *README.md*

file, (3) a *GUI.m* file and (4) *GUI.fig* file. The first two files contain the software license and general repository usage guidance, respectively. The third and fourth, are the source code files supporting the standalone graphical user interface that visually guides a user through this same data preprocessing workflow.

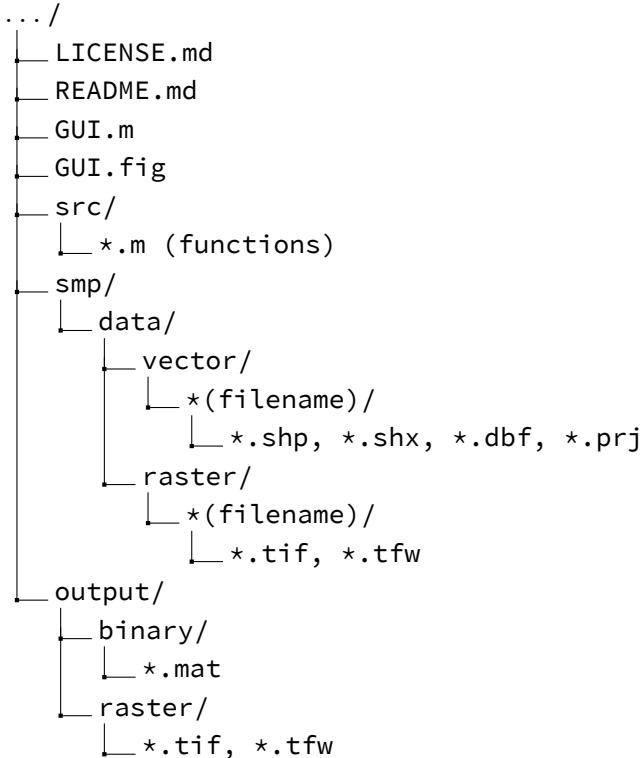


Figure 1.3: Directory tree structure for the toolset repository. Filetypes required for input data and automatically generated as output data are shown.

Also below the top level root directory are the following three subdirectories: (1) *src/*, (2) *smp/*, (3) *output/*. The *src/* directory contains the MATLAB[®] source code m-files comprising the toolset's various functions. The *smp/* directory contains MATLAB[®] .m-file scripts that can optionally be called to automate the execution of multiple data preprocessing workflows. The */smp/data/* directory contains two sub-directories: *vector/* and *raster/*. Each of these houses the corresponding sub-

directories, one for each vector and raster based raw input spatial data files provided by the user. The tiles used for each of these **filename*) sub-directories are automatically assigned to the outputs generated by the tool. The supported vector input filetype is the ESRI shapefile format. Alternatively, the supported raster input file type is the open source GeoTiff format. Finally, the *output*/ directory contains two sub-directories: *binary/* and *raster/*. These subdirectories comprise the default destination locations for all of the outputs generated by the toolset tools. Outputs can be produced in either the mat-file MATLAB® ASCII-binary format or in the same GeoTiff format as the input raster data.

The toolset supports the use of composite raster data sets which are made up of multiple, possibly overlapping, individual raster data tiles. It does this by performing a bounding box intersection test for each input raster data tile with the reference spatial domain. For those tiles whose bounding boxes are found to intersect that of the reference domain, values are iteratively compiled into a new composite mosaic data layer made up of, potentially several, individual tiles. This makes it possible to use input raster data layers that are of arbitrarily high resolution covering large geographic domains.

1.6 AN EXAMPLE IMPLEMENTATION

The raw input datasets which were selected for the example implementation were collected from several publicly available sources. A brief topical description of each source as well as a link to its source web repository is given in Figure 1.4. In addition to these raw input data sources, a number of derived data products are generated automatically from the digital elevation model (DEM) for use in this particular case study analysis. These derived products include: slope & aspect.

Type	Category	Source
Vector	Resource Areas	Cal-Atlas
Vector	County Boundaries	Cal-Atlas
Vector	Surface Geology	Cal-Atlas
Vector	Road Network	Cal-Atlas
Vector	STATSGO Soils	USGS
Vector	State Park Boundaries	Cal-Atlas
Vector	Stream Reaches	National Map
Vector	Street Network	Cal-Atlas
Vector	Surface Water Storage	Cal-Atlas
Raster	Crop Data Layer	USDA
Raster	Digital Elevation Model	National Map
Raster	NLCD Landcover	National Map

Figure 1.4: Table of input data sources used in the case study MCSS model for artificial groundwater recharge applications.

1.7 THE GEOGRAPHIC UNIT OF ANALYSIS

The geographic unit of analysis selected for this example implementation is the US Geologic Survey (USGS) Hydrologic Unit Code (HUC) level five watershed. Specifically, the level five watershed areas contained within the administrative boundaries of the state of California. According to the USGS:

The United States is divided and sub-divided into successively smaller hydrologic units which are classified into four levels: regions, sub-regions, accounting units, and cataloging units. The hydrologic units are arranged or nested within each other, from the largest geographic area (regions) to the smallest geographic area (cataloging units). Each hydrologic unit is identified by a unique HUC consisting of two to twelve digits based on the levels of classification in the hydrologic unit system.²⁷

The level five designation within this HUC framework is comprised of closed contiguous regions possessing an average area of 588 square kilometers. These level five HUC designated areas are often referred to as the HUC-10 watersheds because of their use of a ten digit unique numerical identification code. Within the state of California, there are 1,040 individual HUC-10 watersheds. These watersheds are non-overlapping and have been derived algorithmically from the national elevation dataset by USGS scientists according to the method described by²⁷.

1.8 WOSS MODEL OUTPUTS

Figure 1.5 illustrates a set of sample outputs that were generated by the data preprocessing components of the toolset for an example HUC-10 reference boundary. The reference boundary can either be selected manually, by calling a function which prompts the user to click on map with all of the HUC-10 boundaries drawn on it, or automatically, by specifying the 10-digit code corresponding to the desired HUC-10 watershed. The toolset processes generate an output layer stack – the individual component layers of which are illustrated in the colored inset map panels. Only layers for which there is at least one non-empty data value are included in the generated outputs. Thus, the number of output components may vary depending upon the coverage of the input data layers relative to the domain of the reference boundary. For vector data inputs, the values which are contained in the output are those corresponding to a single attribute field selected by the user. This field must be either a real or coded numeric data type as the raster data format does not support the native representation of categorical variables.

The toolset structure allows for the automated repetition of this data preprocessing workflow for a large number of reference boundaries. In this case study implementation, for example, the toolset was used to prepare a single such output layers stack for each of the 1,040 individual HUC-10 reference boundaries contained within the state of California. With these outputs, a corresponding

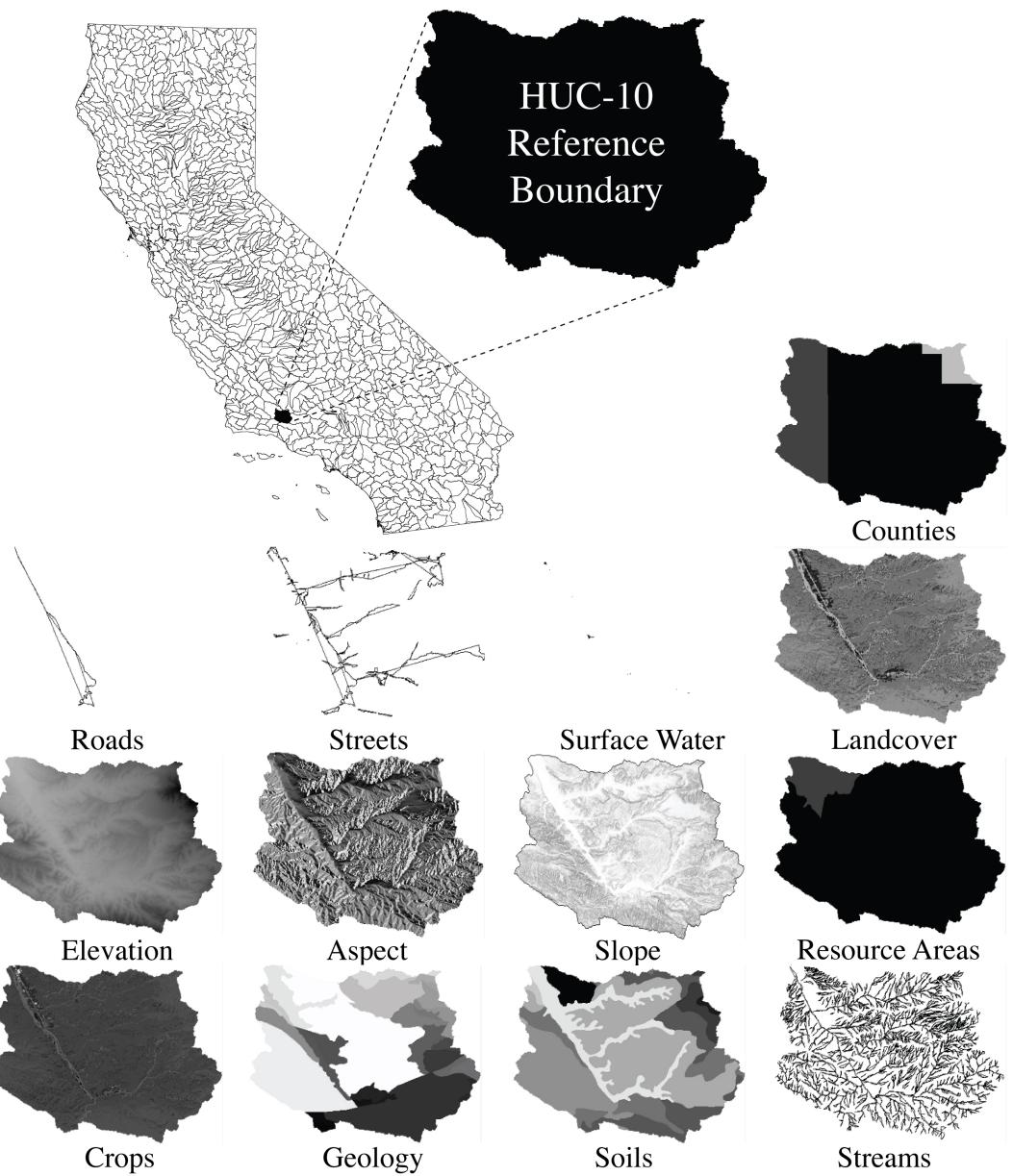


Figure 1.5: Graphical Illustration of the WOSS Model Outputs for Reference HUC-10 Boundary

the MCSS analysis could then be easily conducted for any or every such HUC-10 watershed in the State.

[A computer] takes these very simple-minded instructions – ‘go fetch a number, add it to this number, put the result there, perceive if it’s greater than this other number’ – but executes them at a rate of, let’s say, 1,000,000 per second. At 1,000,000 per second, the results are indistinguishable from magic.

Steve Jobs (1955-2011)

2

Locating Optimal Corridors

2.1 LOCATING OPTIMAL CORRIDORS FOR WATER DISTRIBUTION INFRASTRUCTURE

THE SOURCE OF LOCATION for the delivery of treated wastewater for artificial groundwater recharge applications is always the Wastewater Treatment Plant (WWTP). WWTPs are overwhelming located at low elevation points within the hydrological basins that they serve; as this feature ensures a minimum energy input requirement to deliver wastewater from its distributed sites of generation (i.e. the many homes and businesses distributed geographically throughout the basin) to a single centralized point of treatment. In this way, a system design based upon a single centralized WWTP is best able to take advantage of the use of gravity to provide the motive force for the wastewater's journey through the system. As a side note, an question which may become of substantial future interest is how a more prominent role of reuse may alter our thinking about the optimal design of water treatment collection and conveyance infrastructure. Specifically, it remains an open research question as to whether or not the existing paradigm, with large centrally location treatment facilities, would persist as the most favorable solution if one were charged with designing an integrated wastewater treatment and reuse system from the oft.

The salient output of the first model component - WOSS - is the designation of one or more suitable sites for the placement of either a gravity fed surface spreading infiltration basin or a pump driven subsurface recharge well. In this way, we are left with one or many (in the case of multiple suitable sites) combinations of point source and destination locations. The next modeling challenge which must be overcome therefore is the development of a scheme for generating plausible pathways which connect the source locations to the various destinations. The implicit assumption here being, that in order to deliver the treated wastewater from its source of production, the WWTP, to the end point of use, the recharge site, new water conveyance infrastructure must be constructed.

At this stage it is important to pause to consider whether or not the construction of new water

conveyance infrastructure is in fact a necessary condition of all or any water reuse projects. There are two perspectives from which a response to this objection can be framed. The first is technical in nature. If no new conveyance structure is implemented, two fundamental technical requirements must be met. The first is that the point of reuse must be situated favorably with respect to the existing potable water conveyance system such that it can be readily be connected to, and withdraw from it, large quantities of water for recharge purposes. The second is that the treated wastewater which is to be reused must be returned to a sufficiently high standard of quality such that it can be reincorporated directly back into the potable water supply. This second constraint, while technically feasible given sufficient financial resources, leads naturally to the other category of potential objections to the development of a reuse project without the addition of new conveyance infrastructure: namely, the social stigma associated with co-mingling so-called reclaimed "black water," with the potable freshwater supply. There is a considerable body of research in the social sciences which suggests that a majority of people harbor a very basic, if somewhat irrational, prejudice against the direct reuse of reclaimed water for potable applications.

2.2 THE MULTI-OBJECTIVE CORRIDOR LOCATION PROBLEM

The multi-objective corridor location problem can be formally written as Equation 1³¹. The problem involves the simultaneous minimization of the sums of w independent objective functions \hat{O}_w evaluated at the set of discrete locations x_n comprising a corridor of length n . A valid corridor x_n is subject to the constraint that all of its nodes must be contained within the feasible search domain Ω . Additional, optional constraints which are often imposed upon the structure of x_n shall be discussed in greater detail in subsequent sections.

$$\text{Minimize} : \prod_{n=1}^n \left\{ \hat{O}_1(x_n), \dots, \hat{O}_w(x_n) \right\} \quad (2.1)$$

$$S.t. : \mathbf{x}_n \in \Omega \quad (2.2)$$

Where:

\mathbf{x}_n = The set of discrete row column indices defining a corridor of length n

Ω = The set of discrete row column indices defining the feasible decision space

O_w = The true but unknown forms of w continuous objective functions

\hat{O}_w = The estimates of O_w defined over the discrete set Ω

As a subset of SPPs, corridor location problems tend to be defined in the context of networks with large numbers of nodes and highly structured topologies¹⁴. These shared characteristics arise from the fact that corridor location problems are typically posed in the context of continuous geographic space – a feature which requires that the requisite underlying network structure be generated algorithmically²⁸. In practice, this is frequently accomplished by automating the conversion of a geographically referenced raster grid into a set of nodes by referencing the centroids of the cells within the raster in a process similar to that described by¹⁷. Once the nodes in the network have been created they can then connected to one another by automatically generating arcs using some standard mode of node connectivity; again, using methods similar to those described by⁶.

2.3 GENETIC ALGORITHMS

Genetic Algorithms (GAs) are a family of search heuristics that mimic the process of natural selection to derive one or more near optimal solutions to a given optimization problem.¹² GAs constitute a subset of Evolutionary Algorithms (EAs) which encode solutions using data structures that are analogous to biological chromosomes¹⁰. This feature allows the search for new, better solutions to be accomplished via the iterative application of genetic operations such as crossover, mutation, selec-

tion, etc.¹³ GAs have been developed and profitably used in a wide variety of problem domains from engineering and economics to chemistry and physics. General purpose reviews of the application of GAs to various problems are available from the following references.^{11,33,8} For a more specialized reviews regarding state of the art applications of multi-objective genetic algorithms see the excellent book from Coello & Lamont and, more recently, from Zhou *et al.*^{7,31}

2.4 THE MOGADOR ALGORITHM

MOGADOR is an acronym stands for “Multiple Objective Genetic Algorithm for Corridor Selection Problems.”³⁰ The algorithm was introduced as a novel genetic approach to the problem of multi-objective corridor search.^{22,30} Its development was intended to service need for a robust method of siting optimal corridors relevant to a variety of environmental planning and design applications.^{2,4,32} A need which persisted despite numerous previous efforts to adapt general purpose, deterministic, SSP solution techniques for use in corridor location.^{15,18,19} The continued need for refined algorithmic approaches to the location of optimal corridors within a broad range of application domains is evidenced by the continued appearance of closely related publications during the intervening years.^{1,23,24,25,26,29}

Relative to other traditional shortest path finding routines, the MOGADOR algorithm has been observed to possess a number of favorable characteristics. First among these is the ability of MOGADOR to accommodate large problem statements (large Ω) and/or those which require the simultaneous evaluation of large numbers of independent objectives (large w) (Pangilinan & Janssens 2007, Zhang et al 2008). Another useful feature of the MOGADOR algorithm is its ability to generate an entire solution set from a single run – as opposed to the single solution per run which is typical of traditional SPP algorithms (Cherkassky et al 1996, Dreyfus 1969, Hart et al 1968, Zhang et al 2008). Lastly, but perhaps most importantly, is that when properly parameterized, each of the

solutions generated by the MOGADOR algorithm can be said to be non-inferior to every other (Zhang et al 2008). In this way, the output solution set approximates the so called Pareto optimal solution set for the given problem statement (Deb 2014).

In order to proceed with our discussion of the MOGADOR algorithm, some basic terminology must first be defined. In the context of the MOGADOR algorithm a single gene x is comprised of a pair of row column indices (r, c) to a geographically referenced 2-D array comprising the feasible search domain Ω . An individual I_m is comprised of a sequence of row column index vectors x_n , which collectively form a valid pathway between a predefined set of source $^s x$ and destination $^d x$ locations. In this way, each individual represents a feasible solution to the proposed corridor location problem. A population P_g is comprised of m individuals. And, finally, an evolution E_b is comprised of g populations.

Figure 1 provides a pseudocode description of the MOGADOR algorithm. Its structural components are fairly typical among GAs in general. The search process begins with a stochastic routine for generating of an initial seed population P_1 . Following the initialization of this seed population, the fitness F_1 of each individual in the seed population is computed by summing the objective function scores corresponding to each set of nodes comprising each individual. Upon the completion of this initialization phase, the algorithm then enters a loop wherein successive genetic operations are applied to the initial seed population. In the case of MOGADOR, these operators include: the selection individuals for reproduction on the based upon their fitness, the crossover of selected individuals that share at least one common feature, and finally, the mutation of crossed over individuals so as to maintain a degree of random variation within the population. At the end of each loop iteration a new population is generated, its fitness evaluated, and a convergence parameter computed on the basis of the observed rate of improvement in population fitness across previous iterations. If convergence is achieved ($c_g < ^t c$) the loop is broken and the algorithm terminates returning: the evolutionary history of the population P_{rg} , the corresponding fitness values each individual within

the population $F_{r:g}$, and the convergence parameter history $C_{r:g}$.

Algorithm 1

```

1: procedure MOGADOR
2:    $g = 1 \leftarrow$  initialize loop iterator
3:    $c_g = 0 \leftarrow$  initialize convergence parameter
4:    $P_g = initializeSeedPopulation(m, \Omega)$ 
5:    $F_g = computePopulationFitness(P_g, \hat{O}_w)$ 
6:   while  $c_g < ^t c$  do:
7:      $g+ = 1 \leftarrow$  update loop iterator
8:      $S_g = selectIndividualsFromPopulation(P_g)$ 
9:      $X_g = crossoverSelectedIndividuals(S_g)$ 
10:     $P_g = mutateCrossoverIndividuals(X_g)$ 
11:     $F_g = computePopulationFitness(P_g)$ 
12:     $c_g = computeConvergenceParameter(F_{r:g})$ 
13:   end
14: end while
15: return:  $P_{r:g}, F_{r:g}, C_{r:g}$ 
16: end procedure

```

Figure 2.1: MOGADOR Algorithm Pseudocode

2.5 THE MOGADOR DATA STRUCTURE

The optimal data structure for use in concert with the MOGADOR algorithm is a nested list of lists. Such a list based data structure is well suited to this context as there can be a high degree of variability in the number of elements produced by the different stochastic genetic operators.¹ Figure 2 illustrates a small but valid example such a nested list of lists data structure as used in the context of the MOGADOR algorithm. Note, for example, how the number of row column index vectors n in the first individual I_1 equals 5, while the number of index vectors in subsequent individuals $I_{2:4}$ belonging to the same population, ranges between 3 and 6. The source of this variation has to do with

the stochasticity inherent to the population initialization routine. Similarly, note how the number of populations g in the first evolution E_i is 3 while the number of populations in the second evolution is 4. The source of this variation has to do with the stochasticity in the crossover and mutation processes. Indeed, the only level of the data structure's hierarchy at which a constant number of elements is required is at the population level. Meaning, in other words, that each population P_g contained within evolutions E_b must all possess the same number of m individuals. This requirement ensures that the behavior of the genetic operators is consistent between separate evolutions.

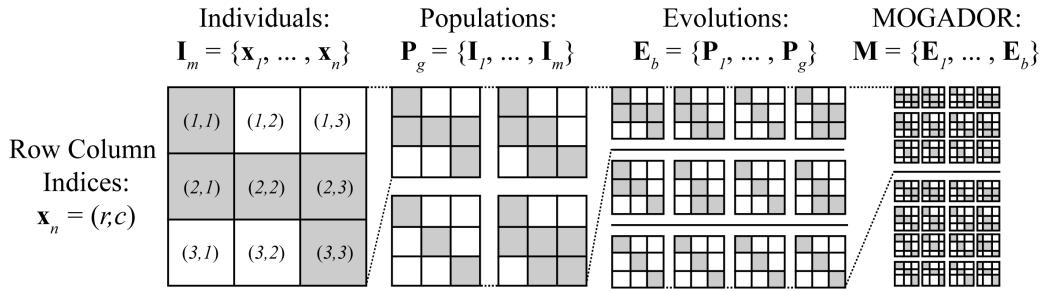


Figure 2.2: MOGADOR Algorithm Data Structure

2.6 INITIALIZING THE MOGADOR ALGORITHM

A novel population initialization procedure has been developed for use in conjunction with the MOGADOR algorithm which improves the global quality of the output solution set while simultaneously reducing overall computational effort. At its core, this novel pseudo-random walk algorithm works by repeatedly sampling a dynamically parameterized bivariate Gaussian distribution. The generic form of the probability density function for the bivariate Gaussian distribution can be written as Equation 2 (Johnson et al 2002). Here, the bivariate Gaussian PDF $f(\tau)$ is function of two inputs. [1] The first is a mean vector μ , comprised of the means (μ_1, μ_2) of two continuous random

variables (τ_1, τ_2) . [2] The second is a covariance matrix Σ , comprised of the pairwise covariances σ for all possible combinations of the continuous random variables (τ_1, τ_2) .

$$f(\tau) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} e^{\frac{-1}{2}(\tau - \mu)^T \Sigma^{-1} (\tau - \mu)} \quad (2.3)$$

Where:

τ = The set of correlated continuous random variables (τ_1, τ_2)

μ = The set of mean values (μ_1, μ_2) for (τ_1, τ_2)

Σ = The pairwise covariance structure $\begin{bmatrix} (\cdot_{11}, \cdot_{12}) & (\cdot_{21}, \cdot_{22}) \\ (\cdot_{21}, \cdot_{11}) & (\cdot_{22}, \cdot_{12}) \end{bmatrix}$ for (τ_1, τ_2)

Each sampled value for τ_1, τ_2 can be reduced to a unit vector and interpreted as a set of row column index deltas. The repeated sampling of the distribution therefore provides a simple yet powerful technique for generating randomized positional changes within a 2-D lattice. Additionally, as shall be discussed in the subsequent sections, the ability to dynamically adjust the parameters of the bivariate Gaussian PDF at any time during the sampling process provides a mechanism by which one is able to functionally constrain the randomness of the walk; hence the term: pseudo-random walk.

Figure 3 provides a pseudocode representation of the proposed pseudo-random walk procedure. Structurally, the routine consists of two nested while loops. At each iteration b of the outer loop a single step x_n along an individual walk I_m is taken. This loop continues until the location of the current step is equal to that of the destination ${}^d x$. At each iteration u of the inner loop a candidate next step Δx_u is produced by random sampling the parameterized bivariate Gaussian PDF $f(x_u)$. Candidate next steps are only considered valid if they are contained within the current valid connected set V_n . The current valid connected set is comprised of all neighboring nodes that not been previously visited and that are inclusive to the search domain Ω . If the candidate step is found to be valid the

inner loop terminates, the outer loop iterates, and the walk process continues. If the valid set is ever found to be empty the outer loop iterator i is reset and the entire process is restarted.

Algorithm 2

```

1: procedure PSEUDO-RANDOM WALK
2:    $n = 1 \leftarrow$  initialize outer loop iterator
3:    $x_n = {}^s x \leftarrow$  initialize individual at source
4:    $I^* = computeEuclideanShortestPath({}^s x, {}^d x)$ 
5:   while  $x_n \neq {}^d x$  do:
6:      $V_n = computeValidConnectedSet(x_{1:n}, \Omega)$ 
7:     if  $V_n = \emptyset$  then:
8:        $n = 1 \leftarrow$  reset outer loop iterator
9:       continue
10:      end if
11:       $u = 1 \leftarrow$  initialize inner loop iterator
12:       $\mu_n = computeOrientationVector(x_n, {}^d x)$ 
13:       $d_n = computeMinimumBasisDistance(x_n, I^*)$ 
14:       $n = n + 1 \leftarrow$  update outer loop iterator
15:      while  $x_n \notin V_n$  do:
16:         $\Sigma_u = computeCovarianceMatrix(d_n, u)$ 
17:         $\Delta x_u = sampleBivariateGaussianDistribution(\mu_n, \Sigma_u)$ 
18:         $x_n = x_{n-1} + \Delta x_u$ 
19:         $u = u + 1 \leftarrow$  update inner loop iterator
20:      end while
21:    end while
22:    return:  $x_{1:n}$ 
23: end procedure

```

Figure 2.3: Pseudo-Random Walk Algorithm Pseudocode

In the process of sampling the parameterized bivariate Gaussian distribution the following three pieces of information are used to functionally constrain the probabilities associated with each candidate next step Δx_u . [1] At the start of each walk the set of array indices corresponding to the Euclidean shortest path I^* from the source to the destination are generated using Bresenham's line

algorithm (Bresenham 1977). Using this set of indices the minimum distance d_n from the current position to the nearest point along this Euclidean shortest path is determined. [2] Next, at each step an orientation vector is computed indicating the orientation of the destination location relative to the current position. [3] Finally, each time the bivariate Gaussian PDF is sampled a counter variable u is iterated.

The orientation vector is interpreted as the mean of the bivariate Gaussian $PDF(\mu_n)$. Alternatively, the minimum distance and iteration variables (d_n, u) are processed as inputs to a generator function which produces the covariance matrix Σ_u . This composition of this generator function ensures that the degree of randomness inherent to the selection of each next step is directly related to number of iterations while at the same time being inversely related to the minimum distance from the current location to the Euclidean shortest path.

2.7 AN EXAMPLE PSEUDO-RANDOM WALK

In an effort to make this pseudo-random walk procedure more comprehensible, particularly with regards to the parameterization of the bivariate Gaussian PDF $f(x_u)$, an example implementation is provided in Figure 4. On the far left of Figure 4 the current status of an arbitrary pseudo-random walk is shown midway through completion. Also drawn, as a broken line, is an abstract representation of the Euclidean shortest path I^* connecting the source location ${}^s x$ to the destination location ${}^d x$. Show at bottom is the current value of the distance parameter $d_n \approx 7.1$. Just to the right of this, in the zoom inset area, the current valid connected set V_n is drawn with the previously visited indices (x_n, x_{n-1}) greyed out to illustrate their elimination from consideration as valid next steps in the walk process. Also shown in this inset are the row column unit vector deltas Δx_u associated with movement to each of the seven nodes contained within the current valid connected set. The small arrow pointing downwards and to the right depicts the current state of the orientation vector μ_n which

describes the position of the destination location d_x relative to the current walk location x_n . The current value of this vector ($\mu_n = [I, I]$) can be thought of as indicating the row and column deltas associated the next step possible step most directly leading towards the destination.

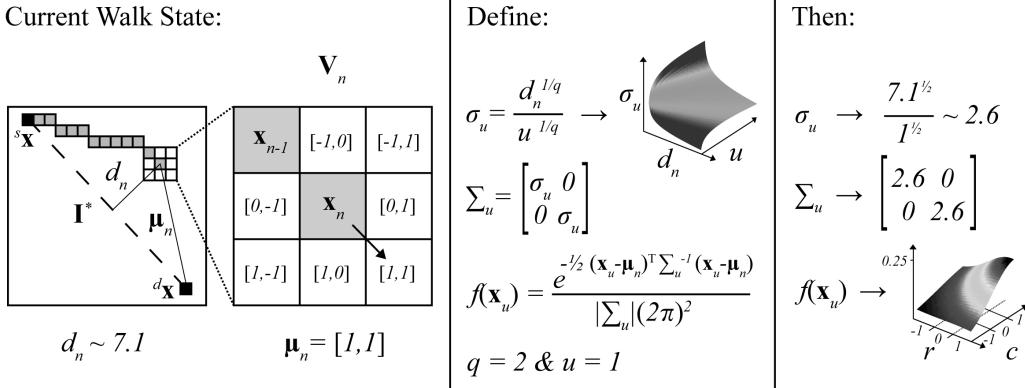


Figure 2.4: Pseudo-Random Walk Example

Continuing on to the right within Figure 4, three functions and two parameter values are defined. [1] First, the covariance term σ_u for the current sample iteration u is specified using a covariance generator function whose form enforces the relationship between distance, iteration count, and covariance previously described. A small demonstration plot of this function's form is provided. [2] Next, the covariance matrix Σ_u is defined by inserting the covariance term σ_u to the diagonal elements of an empty square 2-D matrix. This repeated use of the same covariance term guarantees that for any value of σ_u the output covariance matrix Σ_u will be positive definite. A square, symmetric, and positive definite covariance matrix is a hard requirement for the evaluation of the parameterized bivariate Gaussian PDF $f(x_u)$. The final two variable definitions (q, u) are provided for the sake of computing illustrative values for the other parameters. Numerical evaluations of these expressions are given on the far right portion of the figure. Here again, a small demonstration plot showing the form of the evaluated bivariate Gaussian PDF $f(x_u)$ is provided.

One aspect of this process which warrants further discussion is the role of the fixed parameter q

in determining the degree of randomness exhibited by a given pseudo-random walk. The degree of randomness can be quantitatively defined as the range and extent to which the bivariate Gaussian PDF $f(x_u)$ deviates from its uniform bivariate counterpart. To illustrate this concept consider for example the characteristics of the PDF that would be required to produce a simple random walk using a similarly structured procedure. In such a case the value of $f(x_u)$ would have to be equal for all possible values of Δx_u . Due to the way in which the covariance generator function has been proposed, the q parameter can therefore be used to determine the maximum range of variation in σ_u which can be produced from any combination of (d_n, u) input values. In this way, q does not alter the structure of the bivariate Gaussian PDF $f(x_u)$ but rather only its magnitude. As a result, while the value of q must always be greater than zero to produce real outputs from the covariance generator function, its value is inversely related to the degree of walk randomness.

2.8 INITIALIZING PROBLEMS WITH LARGE DECISIONS SPACES

While the pseudo-random walk procedure can be used to generate an initial seed population for any MOGADOR problem statement; a number of circumstances have been identified in which the performance of the population initialization procedure can be further refined. [1] The first such situation involves problems with extremely large decision spaces – defined as being thousands grid cells or more on a side. [2] The second problem specifications where it is known, a-priori, that the Euclidean path connecting the source to the destination is not entirely feasible.

Historically, corridor location problems which have been posed in the context of extremely large decisions spaces (large Ω) have been considered infeasible both for conventional deterministic SPP optimization techniques as well as for heuristic approaches such as MOGADOR. With regards to MOGADOR, the source of this infeasibility stems from the huge runtime commitment associated with generating and processing populations containing a sufficient number of individuals so as to

ensure that a sufficient amount of genetic diversity can be captured during the initialization phase to conduct a global search.

One strategy which can be employed to ensure enough genetic diversity is produced within the initial seed population without having to generate populations of unduly large size or populations with individuals characterized by a high degree of randomness, is the generation of so called multi-part pseudo random walks. This procedure can be thought of as somewhat analogous to orthogonal statistical sampling techniques such as Latin Hypercube sampling which are used to generate samples from a non-uniformly distributed population by first dividing it into equally probable subspaces (McKay et al 1979, Ye 1998). The implicit assumption here being that the fitness distribution of all possible corridors connecting a typical source and destination pair is similarly non-uniform.

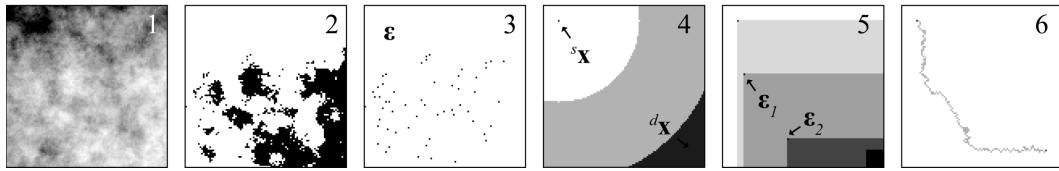


Figure 2.5: Conceptual Illustration of a Multi-Part Pseudo-Random Walk

The multi-part walk process is described by the sequence of panels moving from left to right within Figure 5. [1] The process begins with the far left panel which plots the value of the objective variables within the a square 2-D search domain. [2] The first step is to create a binary mask of feasible nodes by selecting objective surface values less than some arbitrary threshold. [3] The next step involves determining the cell indices for the set of centroid nodes ϵ computed from the connected components within this binary mask. [4] After this, these centroids are assigned rankings on the basis of their inclusion in bins of progressive Euclidean distance from the source location. [5] Next, the procedure requires the iterative selection of centroids ϵ_1, ϵ_2 , one from each successive distance bin, until the bin containing the destination location is reached. The centroid selection process can be unstructured, random within each bin, or structured (as shown), where the centroids considered

eligible for selection each iteration is restricted to those orientated positively in the direction of the destination. [6] Finally, the source is connected to the destination a series of pseudo-random walks constructed for sequential pairs of the selected centroids.

2.9 INITIALIZING PROBLEMS WITH CONCAVE DECISION SPACES

Another circumstance which has the potential to dramatically effect the performance of the pseudo-random walk procedure are problem specifications in which all or a portion of the Euclidean path connecting the source to the destination falls outside the feasible area of the search domain. Such a circumstance can be described as a concave problem, as the source and destination locations are not convex to one another within the boundaries of the decision space Ω . Concave problem statements have the potential to create a situation where at each iteration n of the pseudo-random walk process large values of u must be attained before the covariance term is relaxed enough to allow for the sampling process to generate a Δx_u that is contained within the valid set V_n . In a worst case scenario, the entire runtime improvement associated with the pseudo-random walk based population initialization procedure might be lost.

An approach which has been developed to address such cases is the so called concave multi-part pseudo random walk. It is similar to the standard multi-part walk in that the final walk is composed of a collection of pseudo-random walk sections. However, it differs from the standard walk procedure in that rather than partitioning the space on the basis of distance bands, it iteratively divides the decision space into a series of convex subregions. Here again, these convex sub region contain the centroids associated with connected regions of low objective variable values. The procedure is illustrated conceptually by the sequence of panels contained in Figure 6.

The concave multi-part walk process is described by the sequence of panels moving from left to right within Figure 6. [1] In the first panel, on the far left, we can see a problem that has been

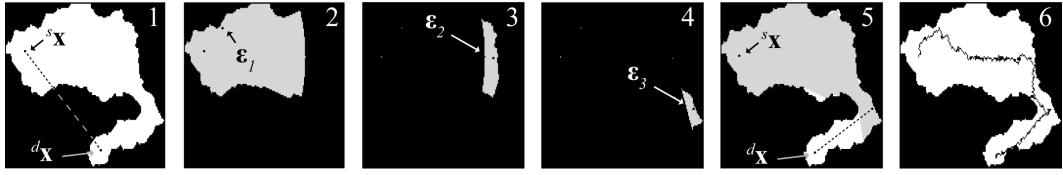


Figure 2.6: Conceptual Illustration of a Convex Multi-Part Pseudo-Random Walk

posed in such a way that the Euclidean path connecting the source to the destination is not feasible as it exits the boundaries of the search domain. [2] For brevity, the two subsequent steps are not shown as they are identical to steps 2-3 in the standard multi-part pseudo random walk procedure. These omitted steps involve the generation of candidate centroids from the connected components within the objective surface. The next illustrated step, shown in the panel second from left, involves computing all of the row column indices that are convex to the source location. Within this convex region ϵ_1 the first centroid ϵ_1 is selected. Here again, the process can be unstructured, where the entire convex region is searched, or structured (as shown), where the initial convex region is restricted to some maximum distance from the source. [3-4] From here, additional non-overlapping convex subregions are computed and candidate centroids iteratively selected from within them. [5] The centroid selection process concludes when the current convex region contains the destination location. [6] Finally, the source is connected to the destination, as in the simple multi-part case, by a series of pseudo-random walks constructed for sequential pairs of the selected centroids.

2.10 MEASURING INITIALIZATION PERFORMANCE

The stochastic processes inherent to the pseudo-random walk procedure, as well as to many other components of the MOGADOR algorithm, make it difficult to analytically derive performance characteristics. As a consequence, with MOGADOR, as with many other GAs, features such as runtime performance must be evaluated through empirical observation. The following sections in-

introduce the results obtained from several such empirical investigations related to the performance of the pseudo-random walk based population initialization procedure for the MOGADOR algorithm. For reference, all computations were performed using a desktop class hardware possessing a 2.3 GHz Intel Quad Core i7 processor (2nd Gen.) with 16 GB of system RAM.

The first of these investigations seeks to understand the role of the fixed parameter q , embedded in the pseudo-random walk covariance generator function, in determining the structural characteristics of output populations. In order to study this issue, a synthetic problem statement was created. This problem statement involves a square search domain of 100 nodes on a side – resulting in a total problem size of $\Omega = 10,000$ nodes. The value of the estimated objective function \hat{O}_w used to evaluate fitness was set as constant for all of the nodes in the search domain. In this way, the objective score is roughly equivalent to individual walk length in Euclidean space. Using this problem statement, twenty seed populations, all containing $m = 100$ individuals, were generated using monotonically increasing values of q beginning with $q = 0.5$ and concluding with $q = 10$. For each the generated populations, the average objective scores for all individuals as well as the standard deviation of the objective scores among individuals were evaluated. The results of this empirical investigation are presented in Figure 7.

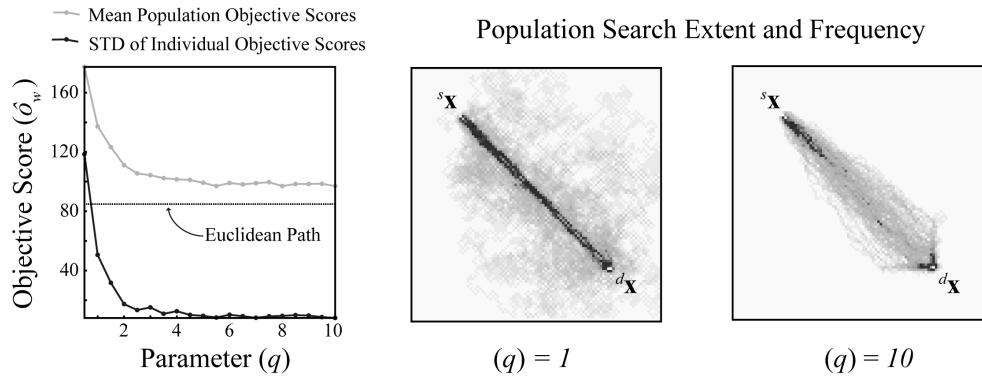


Figure 2.7: Observing the Role of the Fixed Parameter q in Determining Path Randomness

As Figure 7 illustrates, increasing the value of the fixed parameter q causes the individuals within the population to more closely approximate the Euclidean shortest path between the source and destination. Similarly, relaxation of this parameter results in an increase in the perceived randomness among the individual walks within a population. These attributes can be clearly observed in the two images to the right of Figure 7 which show the frequency with which every node within the search domain has been visited by any individual within the two populations generated from different values of q .

Another issue warranting empirical investigation is the relationship between runtime performance of the proposed initialization procedure and problem size. In order to study this issue a series of ten synthetic problem statements were constructed with near identical structural components; differing from one another only in terms of problem size. In each of these problem statements the source location was positioned one fifth of the way down and to the right from the top left of the search domain and, likewise, the destination location was positioned a constant one fifth of the way up from the bottom right corner of the search domain. For all of the different populations generated, the value of the fixed parameter q was set relatively high $q = 10$ to reduce computational effort.

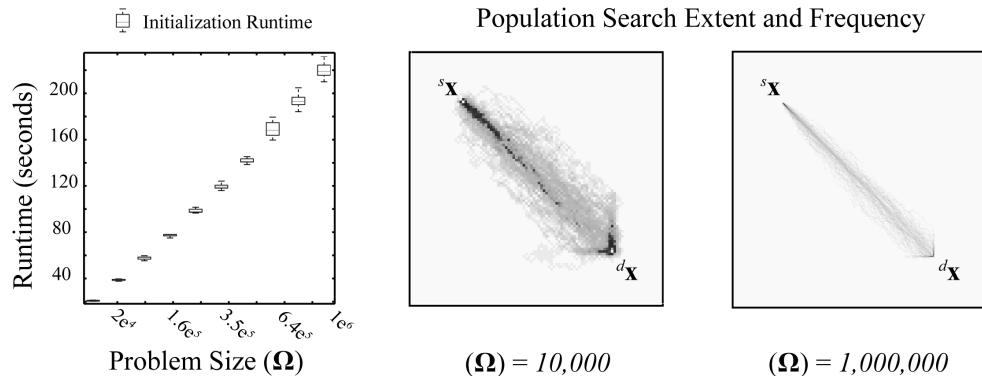


Figure 2.8: Observing the Role of Problem Size on Initialization Runtime

The plot to the left of Figure 8 illustrates the distributional properties of the runtimes re-

quired to generate ten replicate populations for the ten problem statements of progressively increasing size – resulting in a total of $g = 100$ unique populations, each comprised of $m = 100$ individuals. Such repeated simulation is necessary as the proposed initialization routine is based upon a stochastic sampling process which can and will deliver variable runtimes for the repeated applications to the same problem context. One feature of note in this plot is the roughly linear relationship between the mean runtime and problem size for this type of pseudo-random walk based approach to the problem initialization procedure for the range of problem sizes considered. The two images to the right illustrate the effective search extent and frequency for the smallest and the largest populations generated during this investigation.

One of the considerations previously discussed related to the initialization of the MOGADOR algorithm in the context of large problem statements was the need to ensure sufficient diversity within the seed population for the search process to be conducted at the global level. This problem is clearly evident in the population search extent and frequency image contained on the far right of Figure 8. In this example, the population clearly fails to explore a sufficiently large portion of the decision space to be considered as a form of global search. The solution which was previously proposed to this problem involved generating so called multi-part pseudo-random walks. The subsequent investigation therefore, compares the statistical characteristics of a set of populations generated from standard pseudo-random walk to another set of population generated from multi-part pseudo random walks. The results of this investigation are provided in Figure 9.

The runtime reductions which can be achieved from the use of the multi-part pseudo random walk procedure will principally occur in the context of relatively high value settings for the fixed parameter q . This is because while the various component segments of a multi-part walk may not deviate significantly from the Euclidean shortest path however, the randomized connection of multiple such segments produces composite individuals that are significantly more diverse.

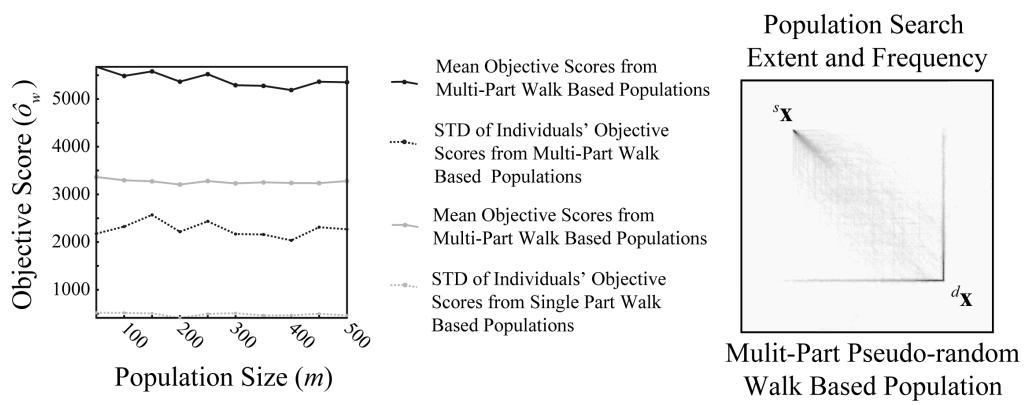


Figure 2.9: Observing the Characteristics of Multi-Part Pseudo-Random Walks

2.II MOGADOR ALGORITHM OUTPUTS

Either the rest of the world can't live like the developed world or we need, as a society, to think more about the technology of providing these services with less intensive use of at least certain resources. We need to do a more diligent job of good housekeeping.

Thomas E. Graedel (1940-)

3

Quantifying Energy-Water Resource Usage

3.1 QUANTIFYING LIFE-CYCLE ENERGY-WATER RESOURCE USAGE EFFICIENCY

3.2 LIFE CYCLE ASSESSMENT AND INVENTORY MODELING

Life-cycle assessment (LCA) is an environmental accounting framework that was developed to systematically quantify the material and energy inputs and outputs from a product, process, or system throughout all its stages of life. It uses a cradle-to-grave or, in some cases, cradle-to-cradle perspective, to evaluate the design processes as well as the entire supply chain associated with manufacturing, transportation, the use phase, and waste management⁷ (Rebitzer2004). Practically, process based LCA analyses, which shall be the focus of the remainder of this discussion, incorporate two distinct modeling phases. The first involves the development of a Life-cycle Inventory model is a cumulative record of all of the materials and energy flows required to deliver a single functional unit of the product or process in question. The second, optional, phase of LCA analyses is the Life-cycle Impact Assessment (LCIA) component. An LCIA model attempts to translate the raw energy and material flows contained within the LCI into different categories of environmental impacts such as global warming potential, ocean acidification, freshwater eutrophication, etc.

In the context of this discussion, one of the stated goals of this dissertation project was to quantify the energy-water usage efficiency of artificial groundwater recharge projects involving the reuse of treated municipal wastewater. In order to accomplish this goal custom LCI models will be developed for five case study regions in which the distribution networks responsible for transporting the treated wastewater for its point of origin, the WWTP, to its destination point of consumption, an artificial groundwater infiltration basin positioned at a designated destination location, upstream within the regional watershed. The raw data supporting the creation of each of these custom LCI models shall be derived from the UC Berkeley supported Web Water-Energy Sustainability Tool (WWEST). In each case study, the scope of this LCI modeling exercise shall be limited to the con-

struction and operational requirements associated with the WWTP plant, the treated water distribution network, and the infiltration basin. This system boundary has been defined in such a way as to emphasize the dynamic contribution of the treated wastewater distribution to the LCI of a given functional unit of treated wastewater delivered to the sub-surface in the face of variable geographic context.

3.3 WASTEWATER TREATMENT PROCESSES

The phrase wastewater treatment encompasses a wide variety of different processes and operational facilities depending upon: the quality of the influent water, the volume of the influent water, and the desired quality/end-use application for the treated effluent water. In the United States the operation of WWTPs are regulated at both the State and Federal levels. At the Federal level the principle regulatory agency is the United States Environmental Protection Agency (USEPA) and the principle regulatory program is the National Pollutant Discharge Elimination System (NPDES). According to the legal mandate of the NPDES, WWTP operators (as well as a wide variety of other entities) are required to apply, at regular time intervals, for discharge permits which provide them with the legal right to release waters containing limited concentrations of regulated pollutants into the environment. Also under this mandate, the USEPA is required to distribute these permits and enforce non-compliance with their terms.

Since the inception of the NPDES program, the USEPA has worked to make readily accessible a centralized database of all registered permit holders within the United States. This database is interesting for the purposes of this project in that it contains spatially referenced information about the operational aspects of every operating WWTP in the U.S. Crucially, this information includes data on maximum daily permitted flow rates and total maximum daily loads that can be used to parameterize the type of process based LCI model facilitated by the WWEST tool.

In terms of their basic physical layout and operational requirements, WWTPs are typically constructed with a tiered layout; comprising primary treatment, secondary treatment, and sometimes, various so called tertiary processes. Both primary and secondary treatment are terms that come with narrow legal definitions and are implemented at nearly all WWTP plants handling municipal sewage discharges. Tertiary treatment processes are more loosely defined and encompass a suite of advanced treatment processes that are so costly that they tend to only be implemented at a minority of WWTP that are subject to a unique circumstances in terms of influent pollutant loadings/composition or requirements associated with designed high purity effluent end-use applications.

Primary treatment encompasses processes and equipment dedicated to the physical separation of non-soluble waste constituents present in the influent wastewater stream. Within a WWTP, a number of distinct processes are often lumped together as being part of the primary treatment. For example, when water first enters the WWTP it is guided through a series of progressively refined grates to screen out bulk pollutants such as anthropogenic trash or natural plant and animal detritus. Following from this bulk screening phase, the water is guided into a series of settling basins where its movement is slowed to crawl to facilitate the settlement of suspended pollutant materials such as sediment. Due to the slow rate at which this settling process proceeds, the physical infrastructure which supports it can comprise a significant fraction of the overall footprint of a WWTP; particularly for those with high flow volume processing requirements.

Secondary treatment encompasses processes and equipment dedicated to the biological (and sometimes chemical) degradation of soluble waste constituents present in the influent wastewater stream. In most municipal WWTP secondary treatment is accomplished through a passively aspirated, aerobic biological digestion reactors. In these reactors large colonies of bacterial species are cultivated on high surface area media using the organic components of the influent wastewater stream as a feedstock for the continued growth. At the end of their life cycle, the bacteria fall to base of the

reactor tank and must be continuously removed in the form of a product known as activated sludge.

Tertiary treatment encompasses processes and equipment dedicated to the removal of soluble inorganic and some organic chemical species (including some viruses and pharmaceutical agents) present within the influent wastewater stream. At present, tertiary treatment processes are not mandatory for all WWTP facilities regulated under the NPDES program. In general, they tend to only be implemented at those specific locations in which a last and credible threat to public or environmental health has been identified and for which a targeted tertiary treatment process exists to address. In this way, mandates for tertiary treatment are typically instigated at the state or local level and done so on a case by case basis. Among the most common tertiary treatment processes include: reverse osmosis filtration, batch irradiation with ultra-violent light, the application of specialized chemical amendments, de-nitrification processes, and others.

For the purposes of this analysis and the customized LCI models which shall be constructed as part of the case study investigations, only primary and secondary treatment processes shall be included in the scope. This decision has been made to eliminate a substantial bias in the inventory models process flows that might be associated with a single specialized tertiary treatment procedure.

3.4 WATER DISTRIBUTION INFRASTRUCTURE

The immediate delivery and reuse of treated wastewater for various municipal and agricultural end-use applications is still a relatively new phenomenon. As such, the regulatory landscape surrounding such practices is still not well defined at the Federal level here in the United States. Consequently, what regulations due exist, typically have been enacted at the State and local levels, with the most advanced frameworks, unsurprisingly, existing in those states such as Florida, California, and Arizona where the popularity of reuse as viable alternative source of freshwater supply, has been surging in recent years.

In all of the locations within the United States for which solid regulatory frameworks surrounding reuse currently exist, there are strong constraints governing the use of existing water distribution infrastructure for the transportation of treated wastewater from its point of origin, the WWTP, to its point of end-use. These regulations, without exception, stipulate that treated wastewater, even if returned to a level of quality consistent with requirements for potable use, cannot be conveyed using existing distribution infrastructure carrying potable water for human consumption in municipal areas. Due to this regulatory constraint, all treated wastewater destined for some sort of municipal reuse must be carried through dedicated parallel distribution infrastructure. In California, this infrastructure is easily identified at locations where treated wastewater is being reused due to the bright purple color of all the pipes. This color encoding is meant to be a strong visual reminder that the water being carried within them has not been deemed, from a regulatory perspective, as being fit for direct human consumption.

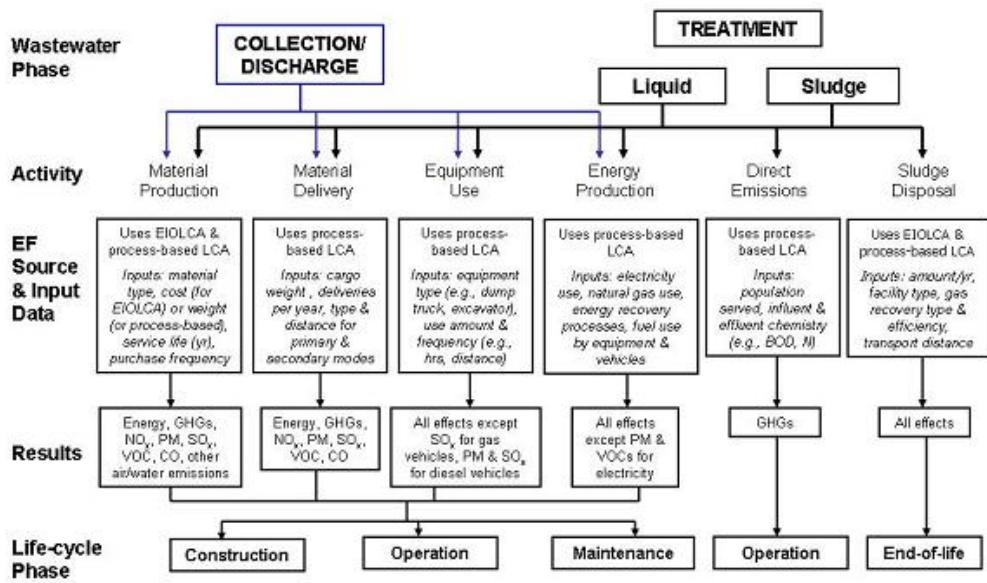
The requirement that treated wastewater, regardless of its standard of treatment and anticipated end use application, be transported using a separate parallel distribution network is expected to be a crucial factor in determining the overall life-cycle energy-water usage efficiency of large scale water reuse systems feeding into artificial groundwater recharge basins. The reasoning behind this expectation is based upon the interaction of the following two key factors. [1] Water is a dense material, and thus it is very energy intensive to transport it over long distances and against steep elevation gradients. [2] Artificial groundwater recharge basins typically require fairly large amounts of contiguous land area that are situated in fairly close proximity to highly developed urban and suburban communities. Municipal water resource management agencies are tightly constrained in terms of the operating budgets from which they are able to draw funds to procure new land holdings for the purpose of constructing artificial recharge basins. Thus, artificial recharge basins, primarily to economic constraints, are typically located fairly far afield from the WWTPs which feed them.

3.5 WWEST RECYCLED WATER REUSE LIFE CYCLE INVENTORY MODEL

The WWEST Recycled Water Reuse Life-cycle Inventory Model refers to an integrated life-cycle inventory database and modeling framework for understanding the environmental impacts of wastewater treatment and reuse processes that was developed by a team of academic researchers operating out of the University of California at Berkeley College of Engineering. The principal investigator behind the project is Dr. Arpad Horvath, Professor of Civil and Environmental Engineering, and a specialist in the environmental impact assessment of civil infrastructure systems. The lead researcher on the WWEST project is Dr. Jennifer Stokes, with the initial development work on the WWEST model being completed as part of her doctoral dissertation project.

The LCI database underlying the WWEST toolset contains process flow information for the manufacture and operation of equipment and facilities involved in the supply, treatment, and distribution of municipal wastewater for the purpose of non-potable reuse. 3.1 provides a schematic overview of the WWEST database model components and their respective input data sources.

The WWEST LCI database can be accessed in two ways. The first, which provides full access to all of the attribute fields contained within the database, is via an excel spreadsheet that has been distributed with a built-in set of computational macros. The second method exposes a more limited range of the data values contained within the database and is accessible via a streamlined web application, also known as WWESTweb. The principle difference between the WWEST excel model and the web application is in the extent to which each allows the user to customize the various parameters associated with a hypothetic water reuse project and its supporting infrastructure. In this sense, the excel based model is more appropriate for building an LCI model to quantify the process flows attributable to an existing facility or which is in a very advanced stage of design planning. Conversely, the WWESTweb framework is more appropriate for less well defined scenario based planning exercises, where the primary goal is to assess the relative impact of major system design



Data can be entered into WWEST in either metric or U.S. units. Some default information about system processes and materials are available in WWEST.

Figure 3.1: WWEST Model System Boundaries

alternatives that have only roughly been fleshed out.

3.6 WWEST MODEL OUTPUTS

In god we trust. All others bring data.

William Edwards Demming (1900-1993)

4

Case Study Implementations

4.1 SANTA BARBARA REGION

4.1.1 REGIONAL CONTEXT

- HUC-8 Code: 18060013
- Total Area: 1,173.6 km^2
- Maximum Elevation: 1,376.7 m
- Minimum Elevation: -0.7 m
- Mean Slope: 13.98 %
- Standard Deviation of Slope: 11.07 %
- Dominant Soil Composition: Hydrologic Soil Group - B: 10 – 20% clay, 50 – 90% sand, 35% rock fragments



Figure 4.1: Santa Barbara Region Overview (Filled in Black)

4.1.2 SEARCH DOMAIN

- Grid Dimensions: 363 cells x 1351 cells
- Grid Cell Resolution: 100 m x 100 m (1 ha)
- Feasible Grid Cells: 117,363 cells

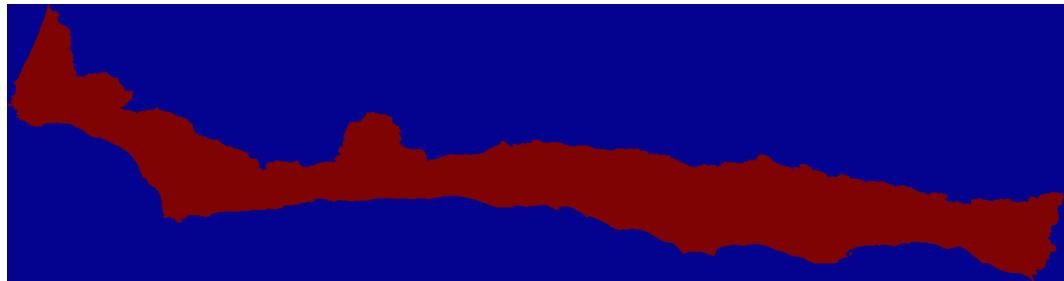


Figure 4.2: Santa Barbara Region Search Domain (Filled in Red)

4.1.3 DESTINATION SEARCH INPUTS

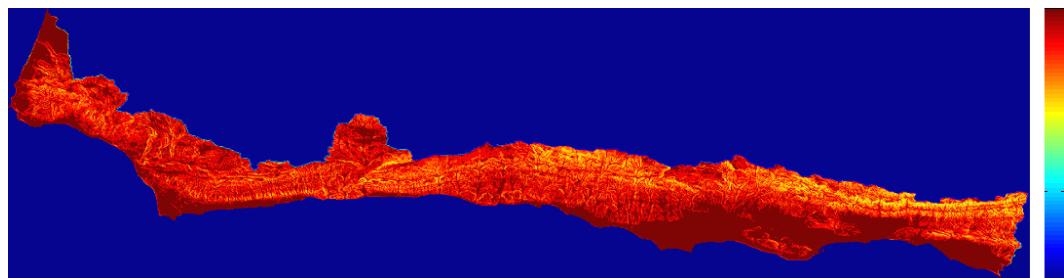


Figure 4.3: Santa Barbara Region Destination Search Inputs: Slope Score (Blue:Low, Red:High)

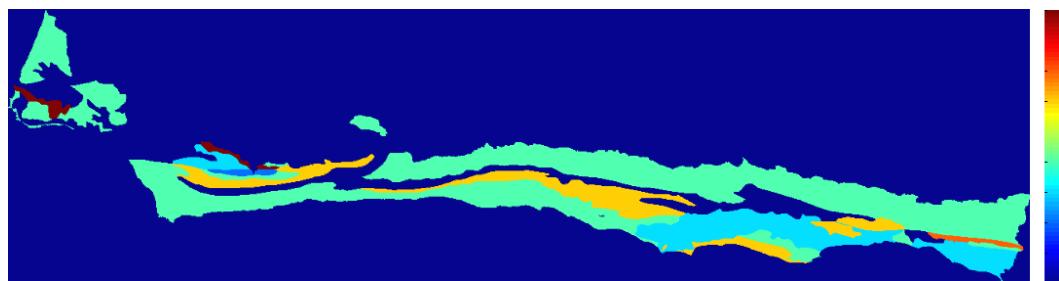


Figure 4.4: Santa Barbara Region Destination Search Inputs: Geology Score (Blue:Low, Red:High)

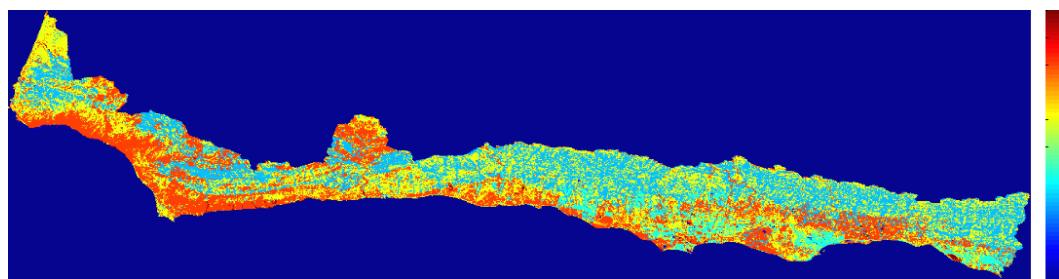


Figure 4.5: Santa Barbara Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High)

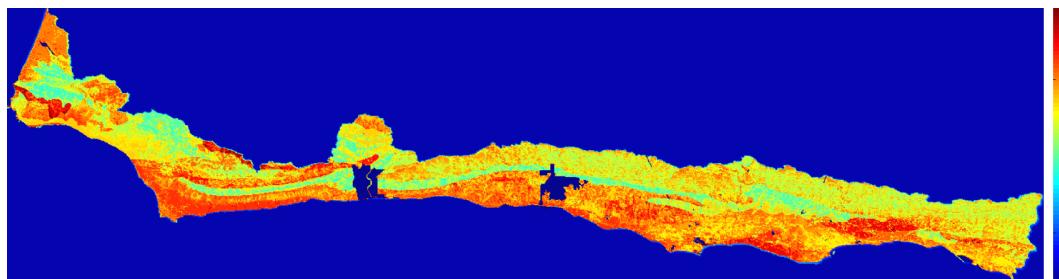


Figure 4.6: Santa Barbara Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)

4.1.4 DESTINATION SEARCH OUTPUTS

4.1.5 PROPOSED CORRIDOR ENDPOINTS

- Start Location: (313, 1083)
- End Destination:



Figure 4.7: Santa Barbara Region Destination Search Outputs: Candidate Regions

- Shortest Euclidean Path Distance:

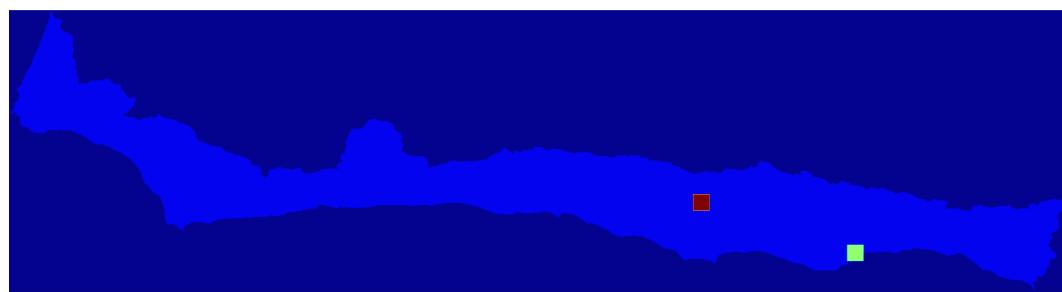


Figure 4.8: Santa Barbara Region Proposed Corridor Endpoints



Figure 4.9: Santa Barbara Region Accessibility Based Objective Scores (Blue:Low, Red:High)

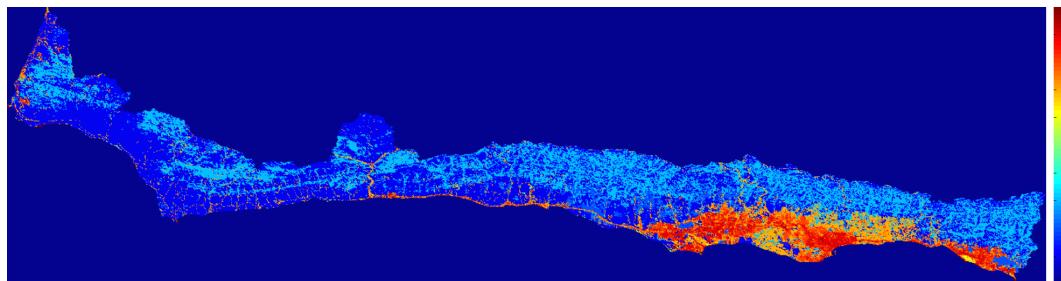


Figure 4.10: Santa Barbara Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)

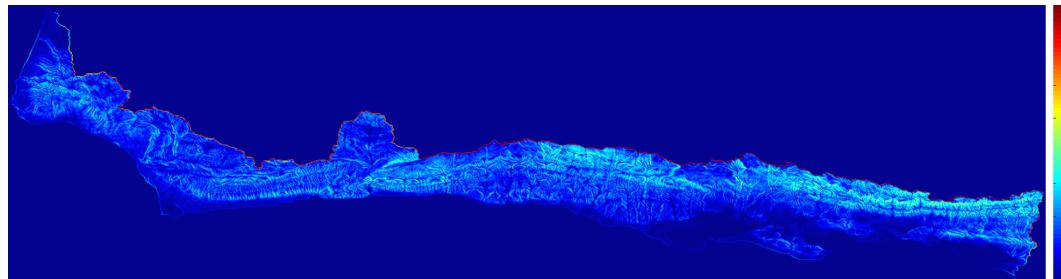


Figure 4.11: Santa Barbara Region Slope Based Objective Scores (Blue:Low, Red:High)

4.1.6 PROPOSED OBJECTIVE LAYERS

4.1.7 PROPOSED CORRIDOR SOLUTIONS

4.1.8 ANTICIPATED DISTRIBUTION OF LIFE CYCLE ENERGY USAGES AND NET WATER SAVINGS

4.2 OXNARD REGION

58

4.2.1 REGIONAL CONTEXT

- HUC-8 Code: 18070102

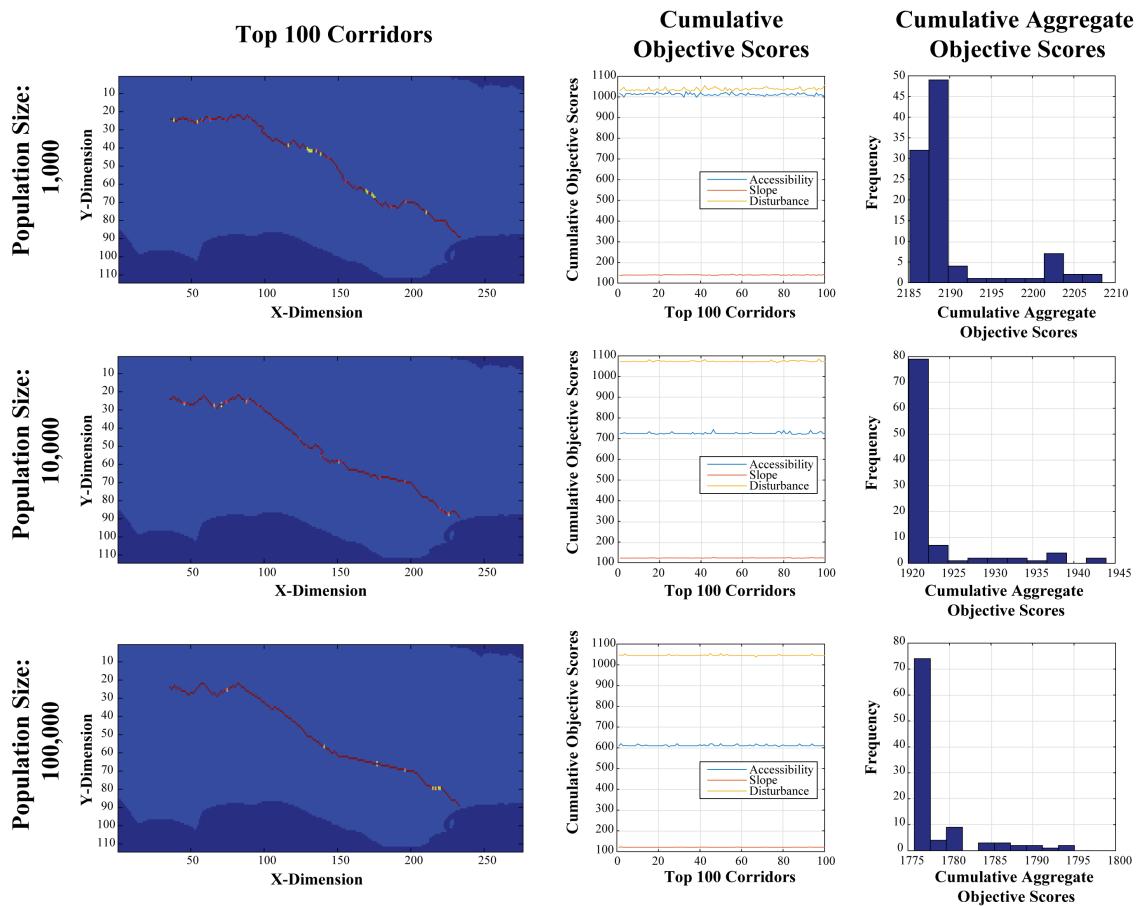


Figure 4.12: Santa Barbara Region Corridor Analysis Results



Figure 4.13: Santa Barbara Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview

- Total Area: $5,188.3 \text{ km}^2$
- Maximum Elevation: $2,664.4 \text{ m}$
- Minimum Elevation: -0.05 m
- Mean Slope: 15.54%
- Standard Deviation of Slope: 11.11%
- Dominant Soil Composition: Hydrologic Soil Group - B: 10 – 20% clay, 50 – 90% sand, 35% rock fragments

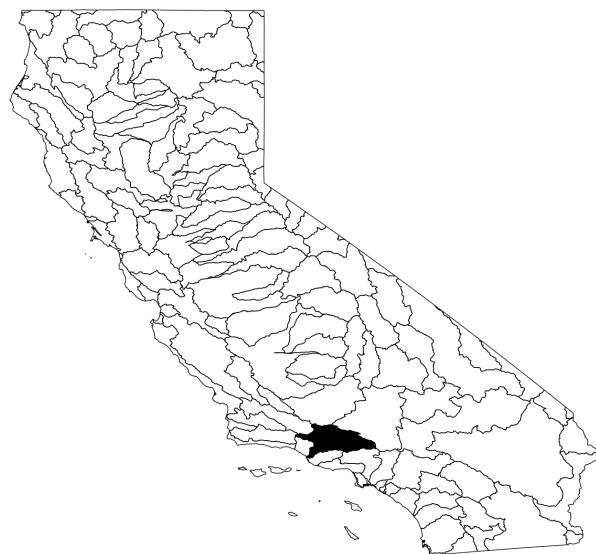


Figure 4.14: Oxnard Region Overview (Filled in Black)

4.2.2 SEARCH DOMAIN

- Grid Dimensions: $677 \text{ cells} \times 1586 \text{ cells}$
- Grid Cell Resolution: $100 \text{ m} \times 100 \text{ m} (1 \text{ ha})$
- Feasible Grid Cells: $518,834 \text{ cells}$

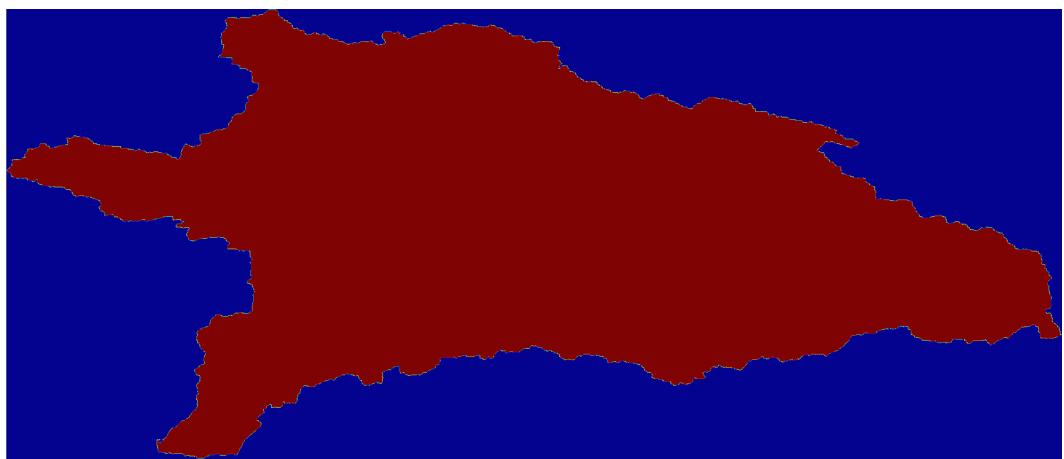


Figure 4.15: Oxnard Region Search Domain (Filled in Red)

4.2.3 DESTINATION SEARCH INPUTS

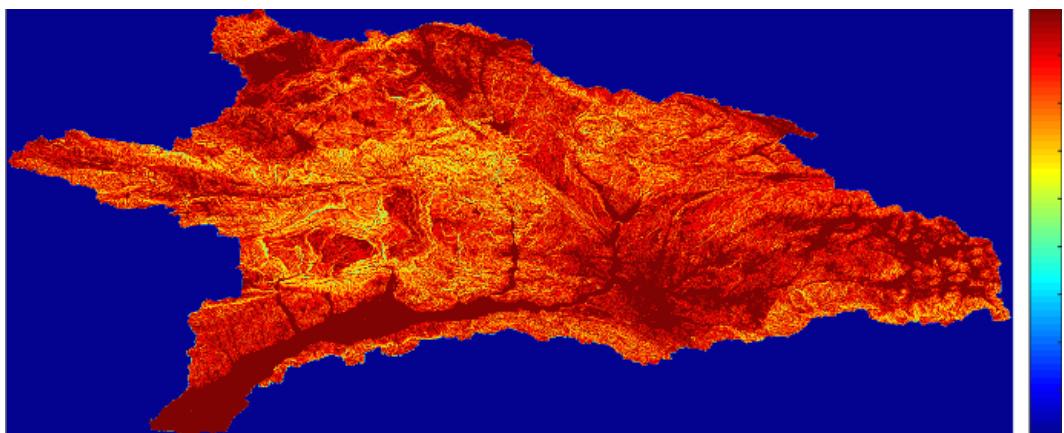


Figure 4.16: Oxnard Region Destination Search Inputs: Slope Score (Blue:Low, Red:High)

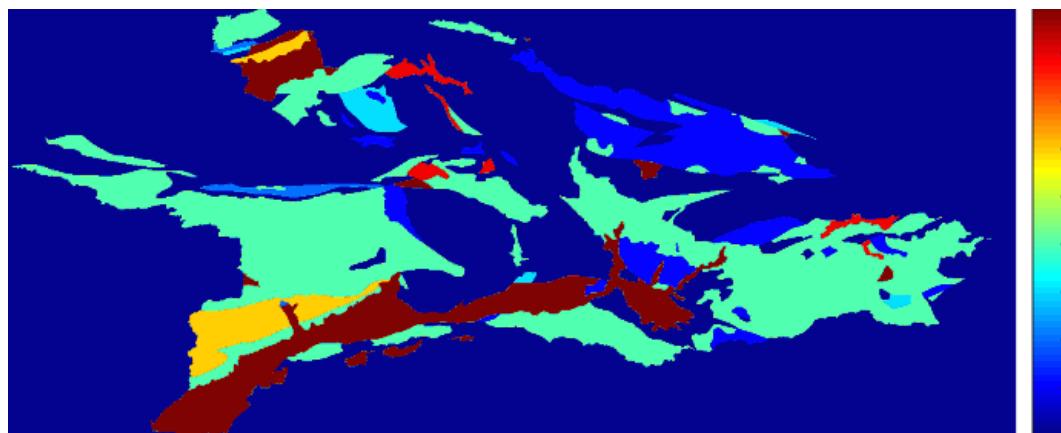


Figure 4.17: Oxnard Region Destination Search Inputs: Geology Score (Blue:Low, Red:High)

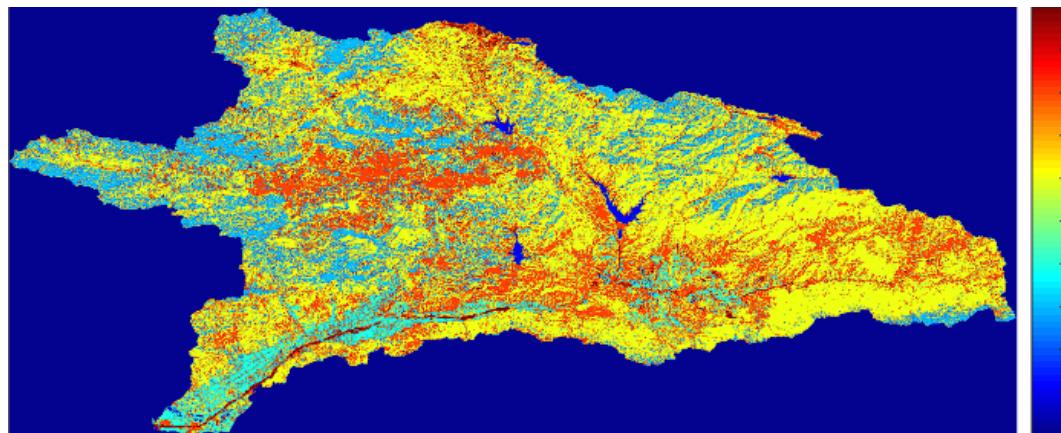


Figure 4.18: Oxnard Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High)

4.2.4 DESTINATION SEARCH OUTPUTS

4.2.5 PROPOSED CORRIDOR ENDPOINTS

- Start Location: (656, 236)
- End Destination: (513, 532)
- Shortest Euclidean Path Distance:

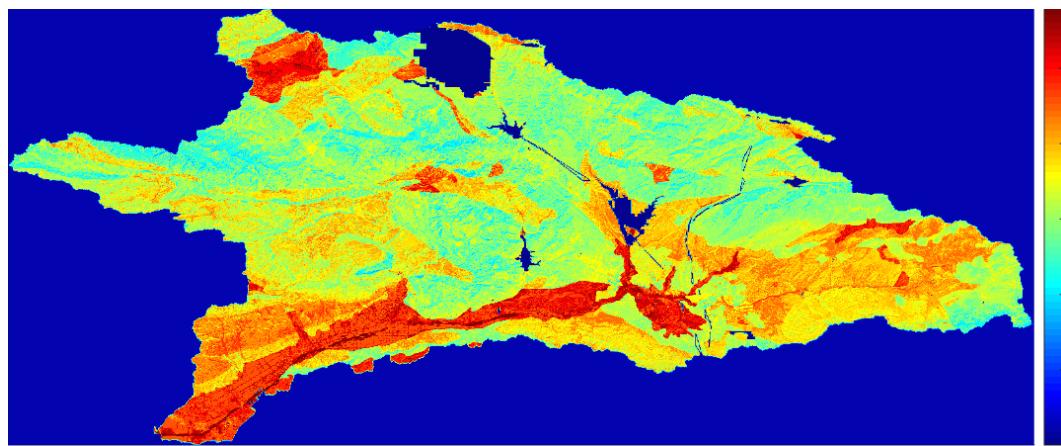


Figure 4.19: Oxnard Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)

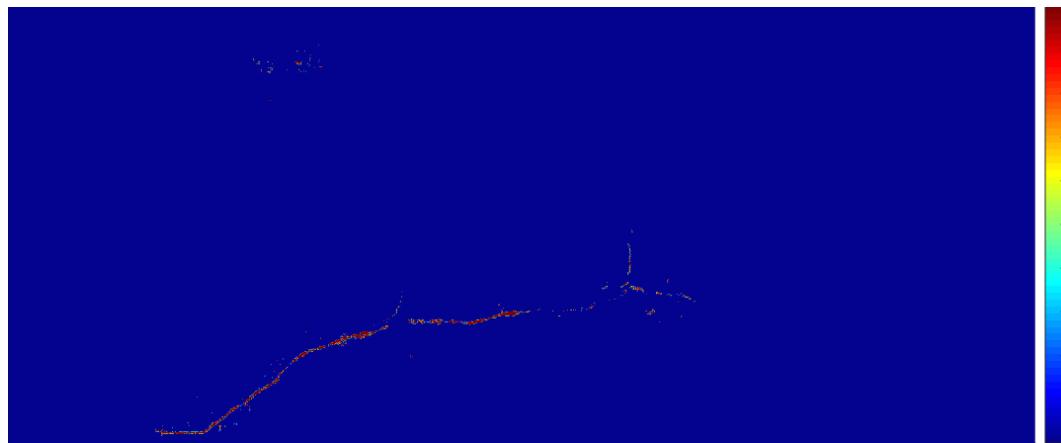


Figure 4.20: Oxnard Region Destination Search Outputs: Candidate Regions

4.2.6 PROPOSED OBJECTIVE LAYERS

4.2.7 PROPOSED CORRIDOR SOLUTIONS

4.2.8 ANTICIPATED DISTRIBUTION OF LIFE CYCLE ENERGY USAGES AND NET WATER SAVINGS

4.3 SAN DIEGO REGION

4.3.1 REGIONAL CONTEXT

63

- HUC-8 Code: 18070304

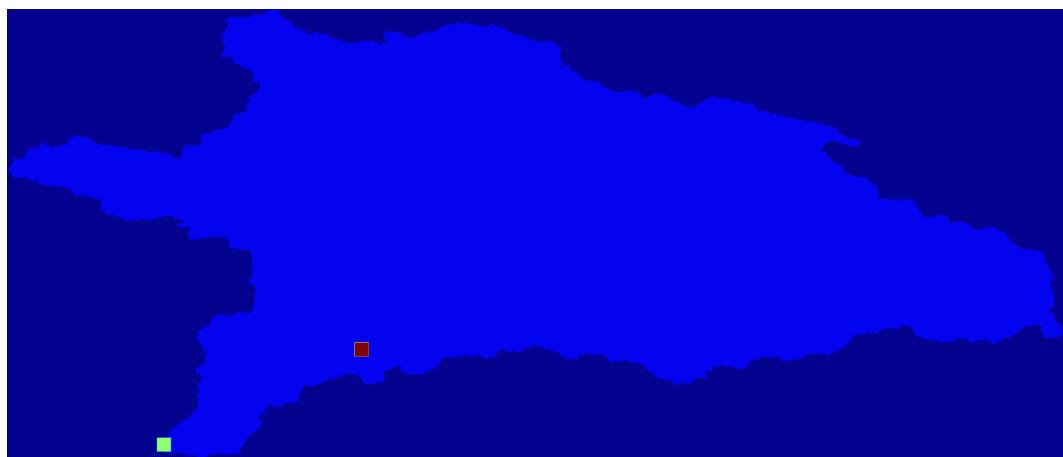


Figure 4.21: Oxnard Region Proposed Corridor Endpoints

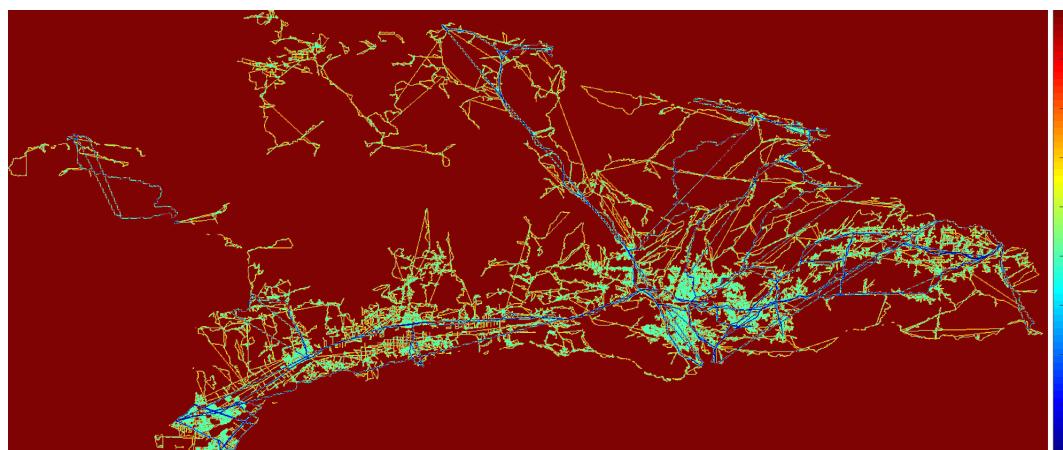


Figure 4.22: Oxnard Region Accessibility Based Objective Scores (Blue:Low, Red:High)

- Total Area: $4,338.1 \text{ km}^2$
- Maximum Elevation: $1,977 \text{ m}$
- Minimum Elevation: -0.7 m
- Mean Slope: 9.38%
- Standard Deviation of Slope: 8.77%

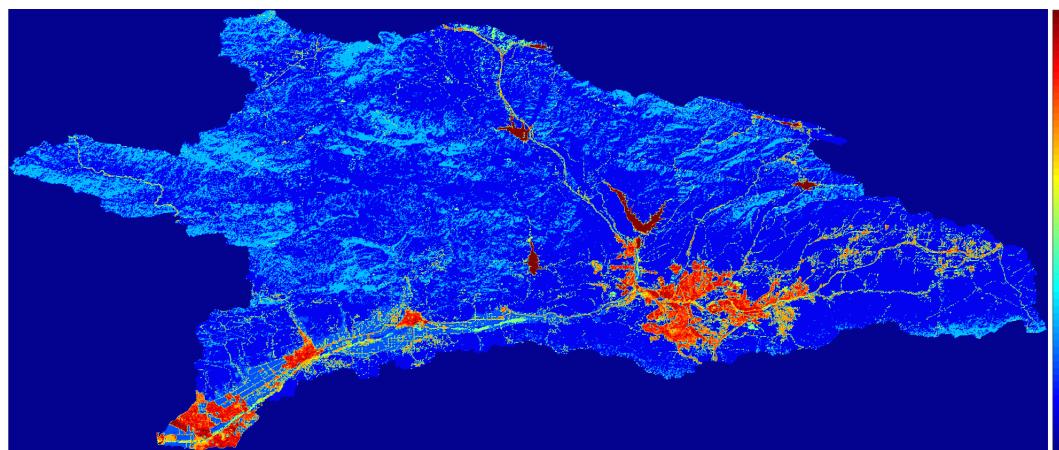


Figure 4.23: Oxnard Region Land Use Based Disturbance Objective Scores (Blue:Low, Red:High)

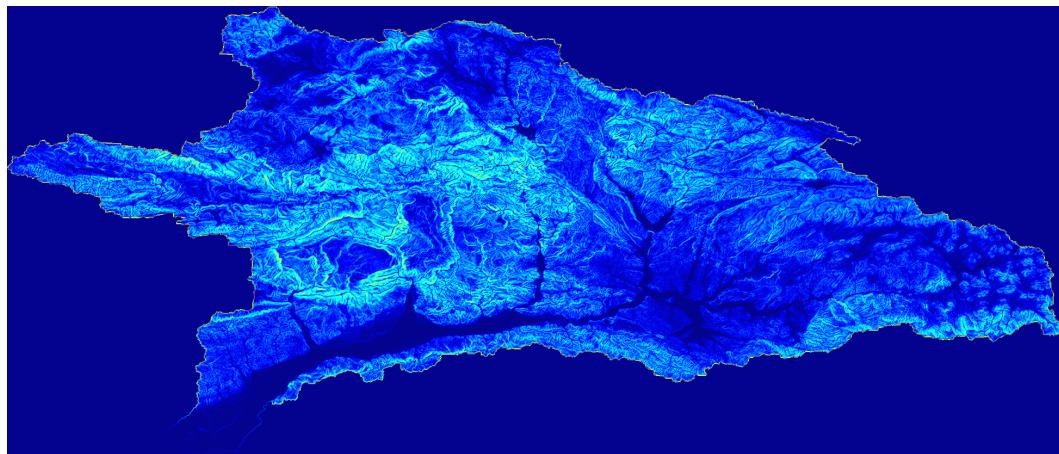


Figure 4.24: Oxnard Region Slope Based Objective Scores (Blue:Low, Red:High)

- Dominant Soil Composition: Hydrologic Soil Group - B: 10 – 20% clay, 50 – 90% sand, 35% rock fragments

4.3.2 SEARCH DOMAIN

- Grid Dimensions: 798 *cells* x 898 *cells*

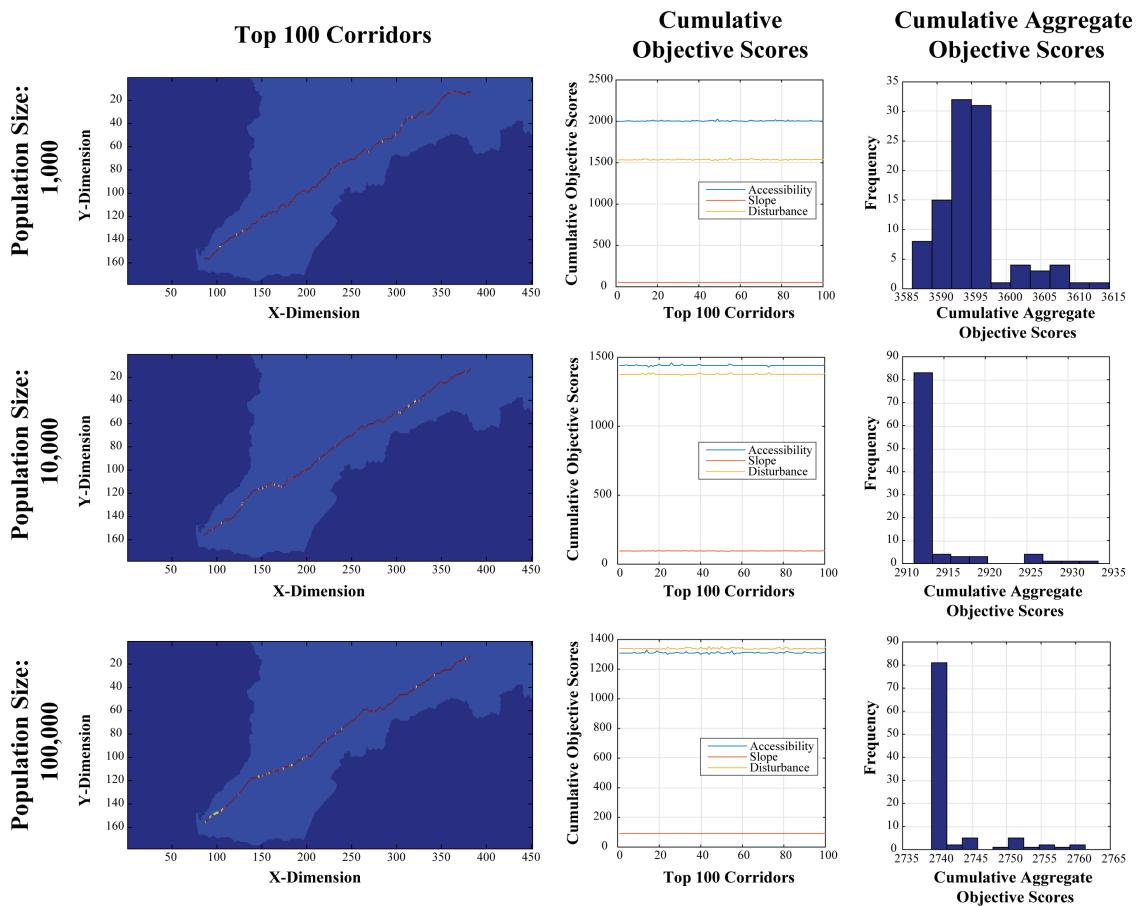


Figure 4.25: Oxnard Region Corridor Analysis Results

- Grid Cell Resolution: 100 m x 100 m (1 ha)
- Feasible Grid Cells: 433, 808 cells

4.3.3 PROPOSED CORRIDOR ENDPOINTS

- Start Location: (635, 42)
- End Destination: (453, 363)
- Shortest Euclidean Path Distance: 36, 900.64 m (36 km)



Figure 4.26: Oxnard Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview

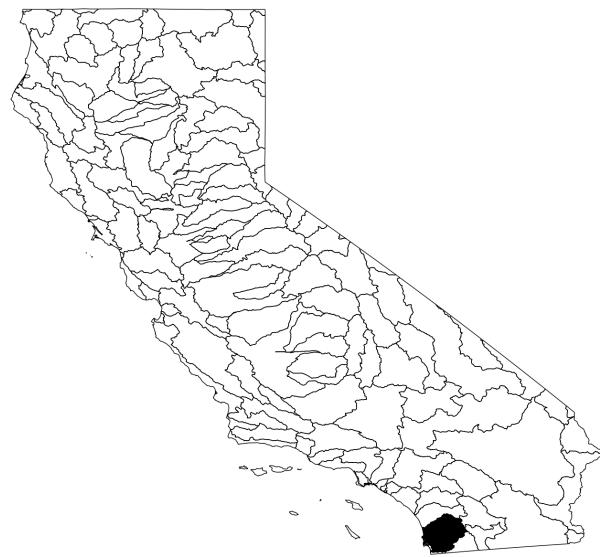


Figure 4.27: San Diego Region Overview (Filled in Black)

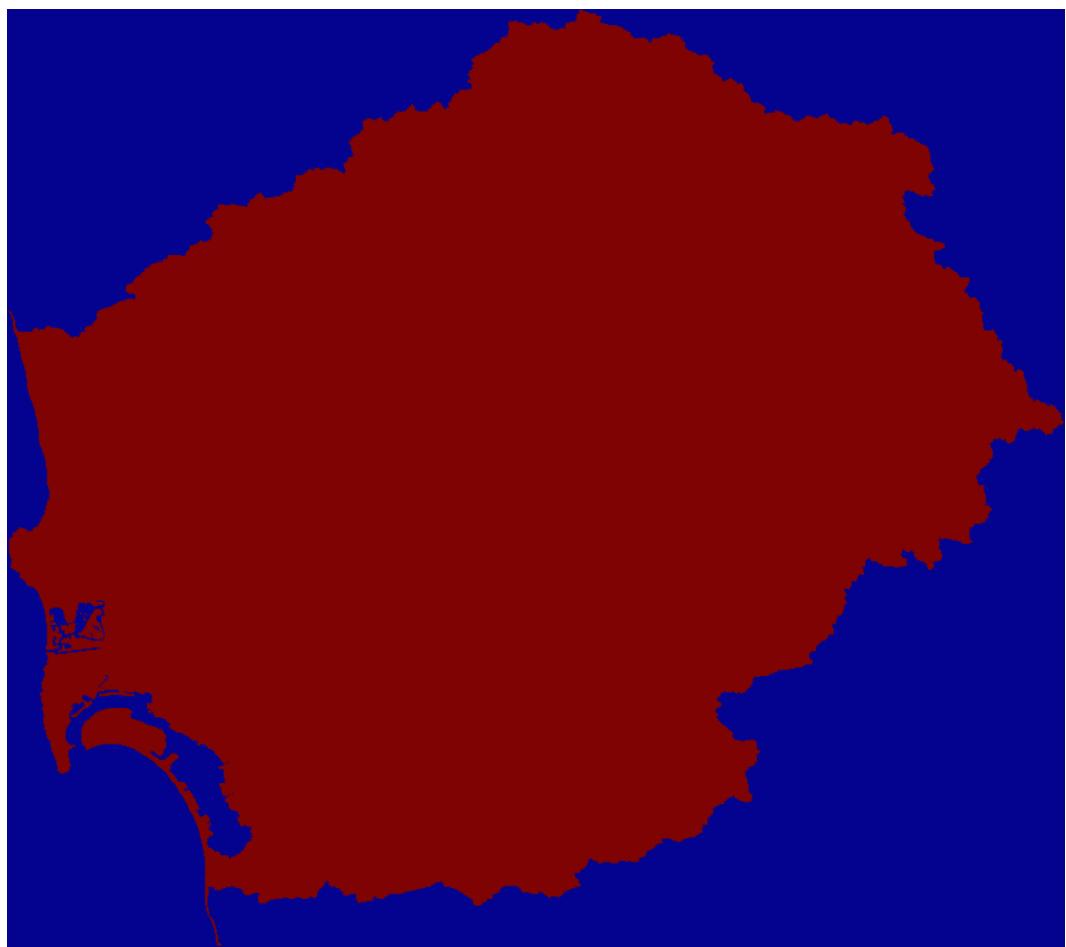


Figure 4.28: San Diego Region Search Domain (Filled in Red)

4.3.4 PROPOSED OBJECTIVE LAYERS

4.3.5 PROPOSED CORRIDOR SOLUTIONS

4.3.6 ANTICIPATED DISTRIBUTION OF LIFE CYCLE ENERGY USAGES AND NET WATER SAVINGS

4.4 SANTA ANA – SAN BERNADINO REGION

4.4.1 REGIONAL CONTEXT

- HUC-8 Code: 18070203

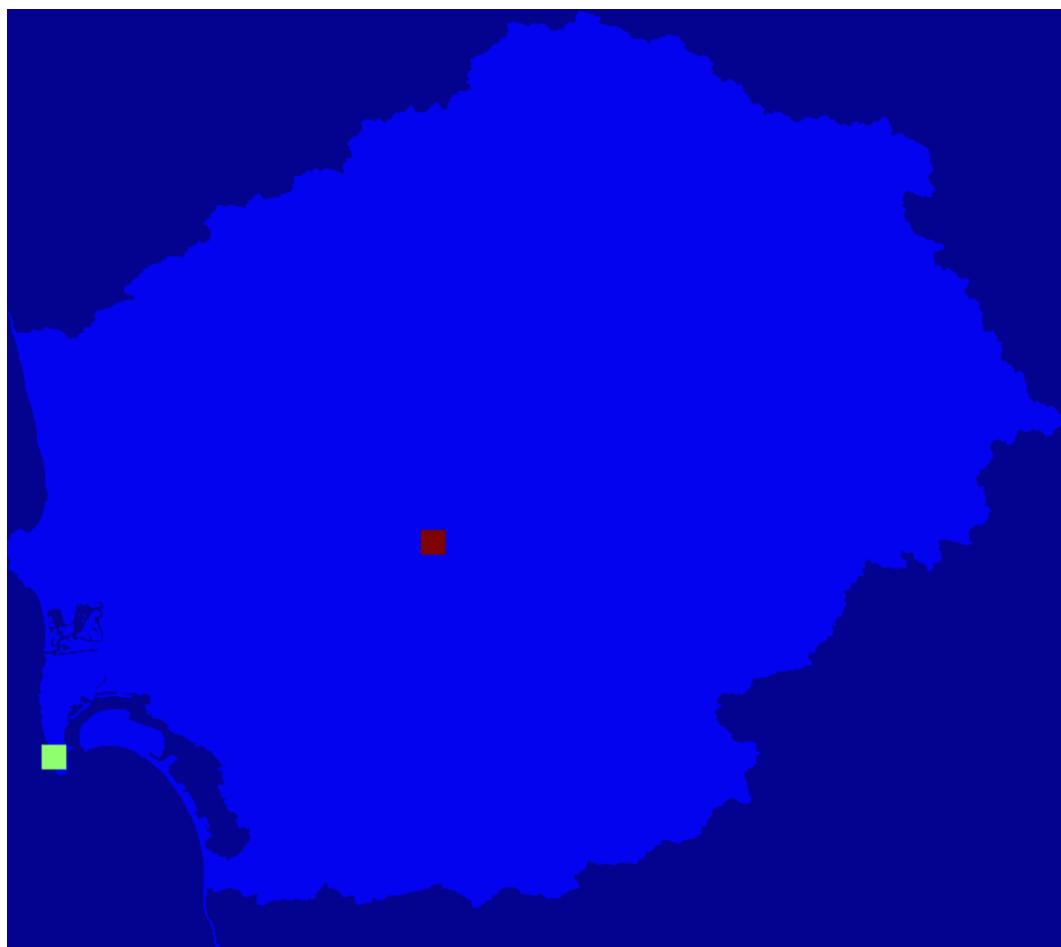


Figure 4.29: San Diego Region Proposed Corridor Endpoints

- Total Area: $5,375.9 \text{ km}^2$
- Maximum Elevation: $3,461.3 \text{ m}$
- Minimum Elevation: -0.7 m
- Mean Slope: 10.56%
- Standard Deviation of Slope: 12.21%
- Dominant Soil Composition: Hydrologic Soil Group - B: $10-20\%$ clay, $50-90\%$ sand, 35% rock fragments

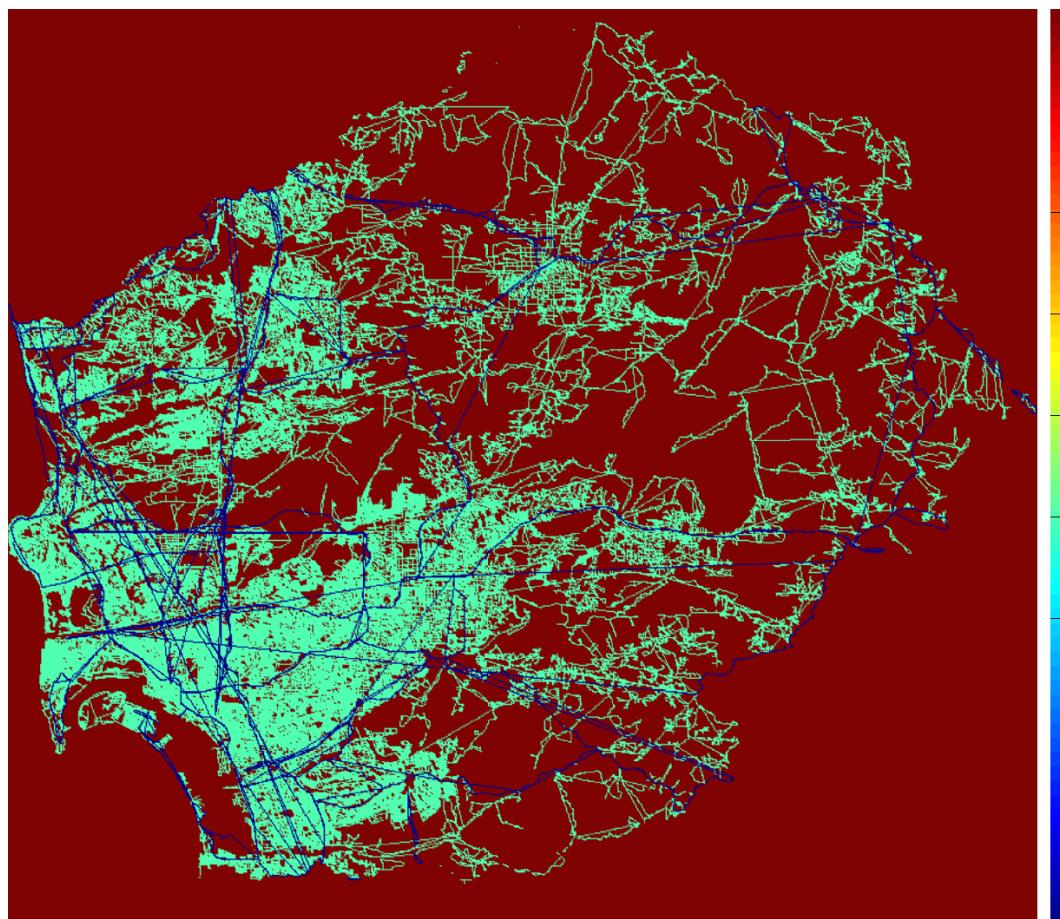


Figure 4.30: San Diego Region Accessibility Based Objective Scores (Blue:Low, Red:High)

4.4.2 SEARCH DOMAIN

- Grid Dimensions: 854 *cells* x 1463 *cells*
- Grid Cell Resolution: 100 m x 100 m (1 ha)
- Feasible Grid Cells: 537,587 *cells*

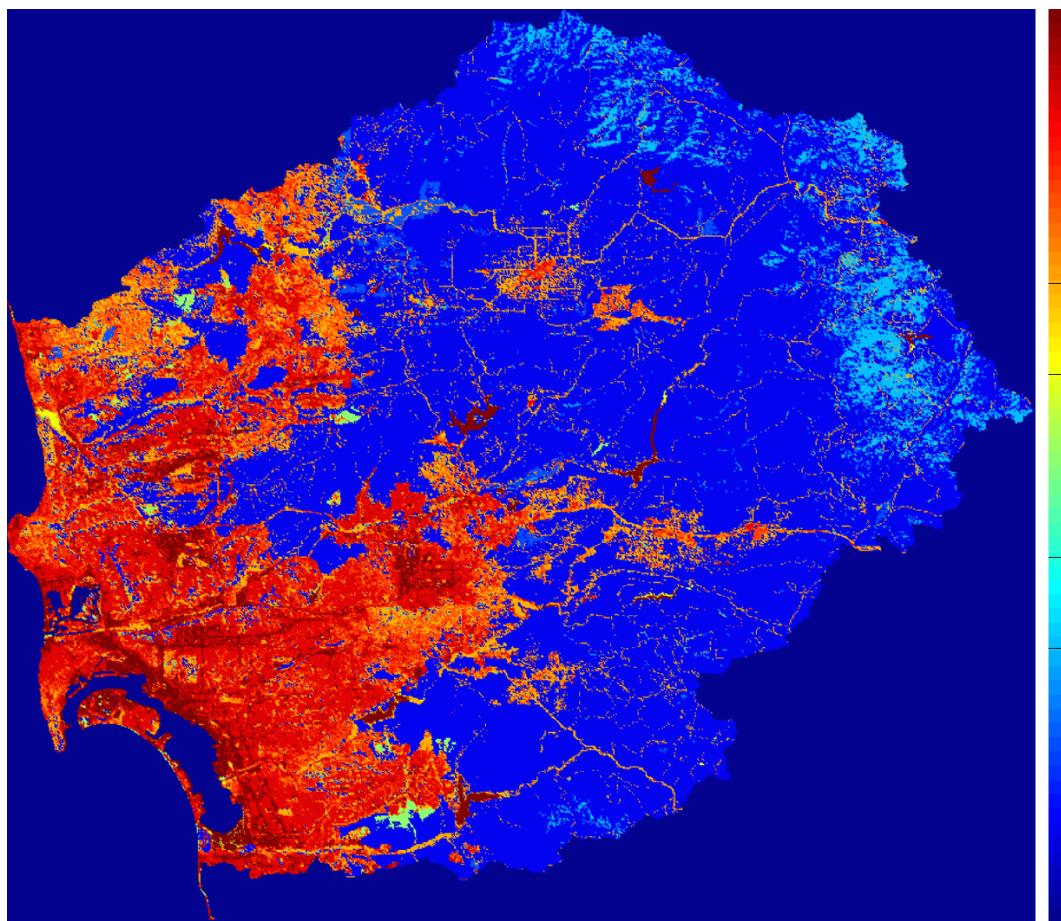


Figure 4.31: San Diego Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)

4.4.3 DESTINATION SEARCH INPUTS

4.4.4 DESTINATION SEARCH OUTPUTS

4.4.5 PROPOSED CORRIDOR ENDPOINTS

- Start Location: (840, 48)
- End Destination:
- Shortest Euclidean Path Distance:

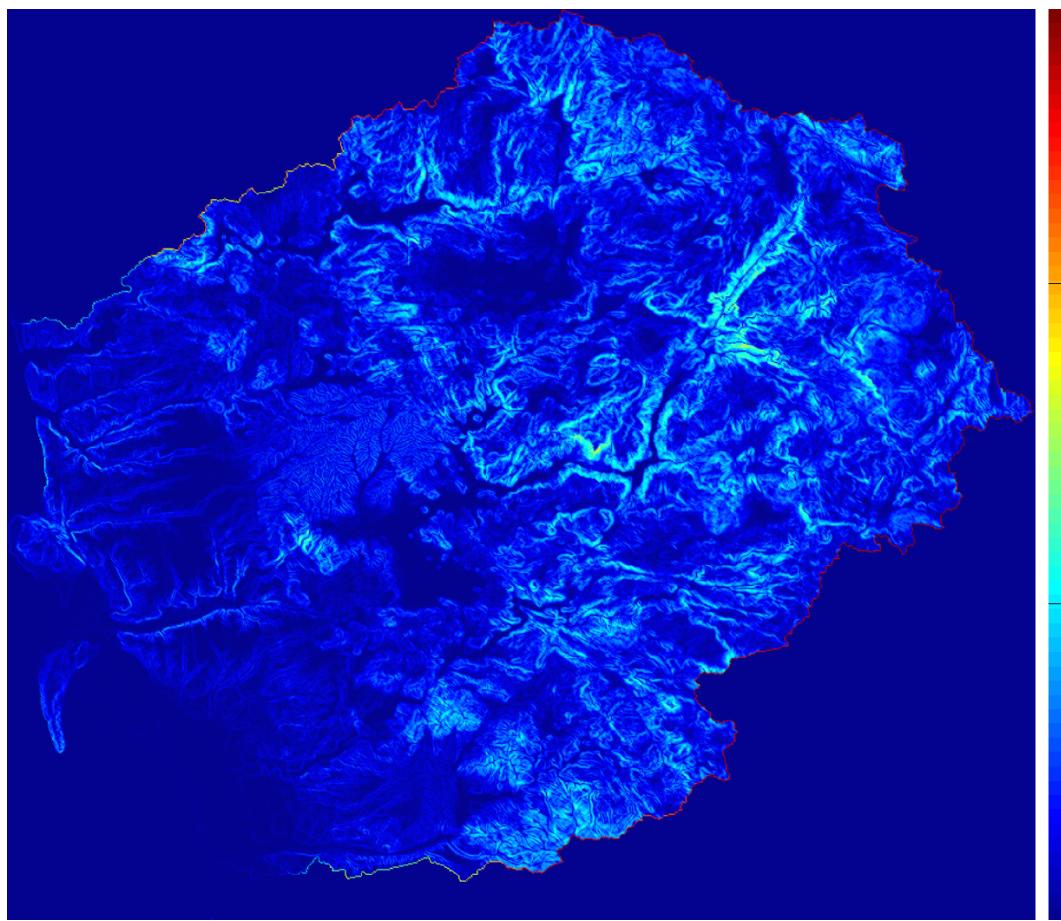


Figure 4.32: San Diego Region Slope Based Objective Scores (Blue:Low, Red:High)

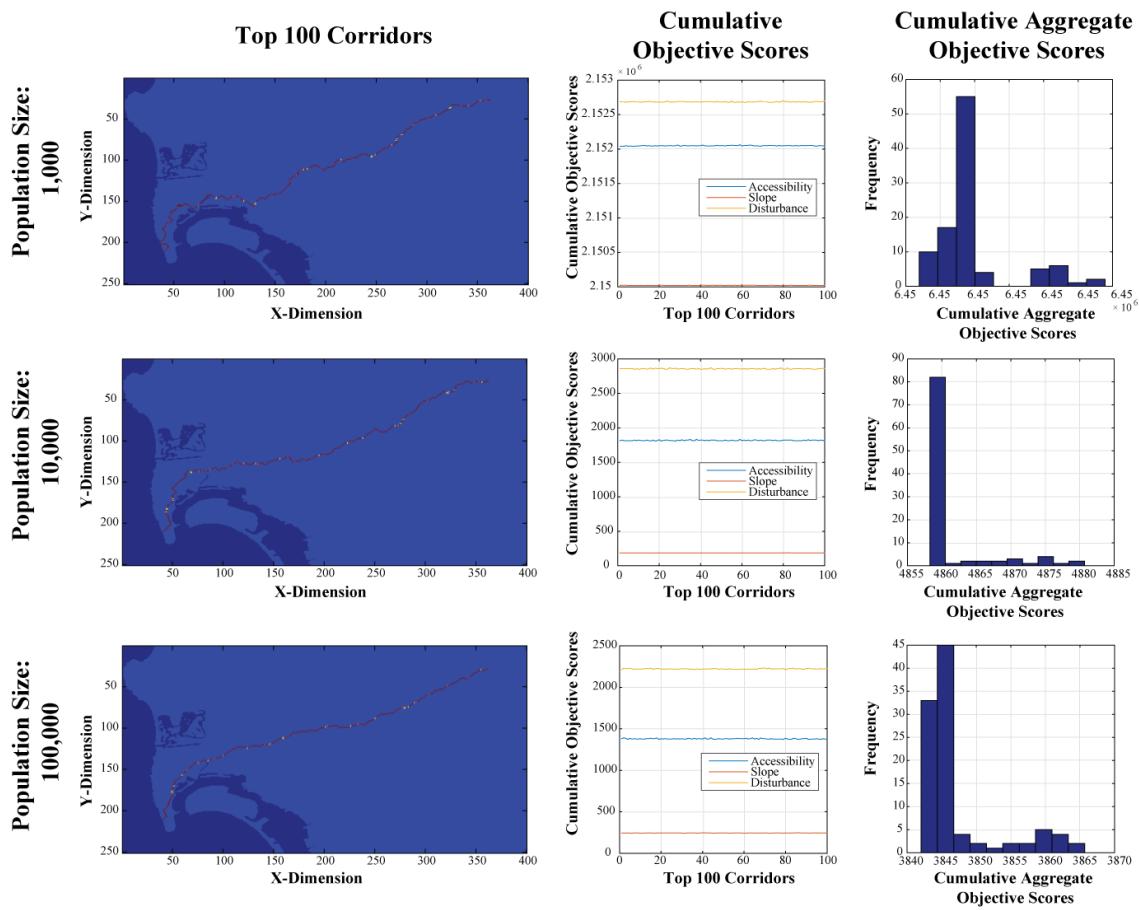


Figure 4.33: San Diego Region Corridor Analysis Results

4.4.6 PROPOSED OBJECTIVE LAYERS

4.4.7 PROPOSED CORRIDOR SOLUTIONS

4.4.8 ANTICIPATED DISTRIBUTION OF LIFE CYCLE ENERGY USAGES AND NET WATER SAVINGS

4.5 FRESNO – TULARE REGION

4.5.1 REGIONAL CONTEXT

- HUC-8 Code: 18030009

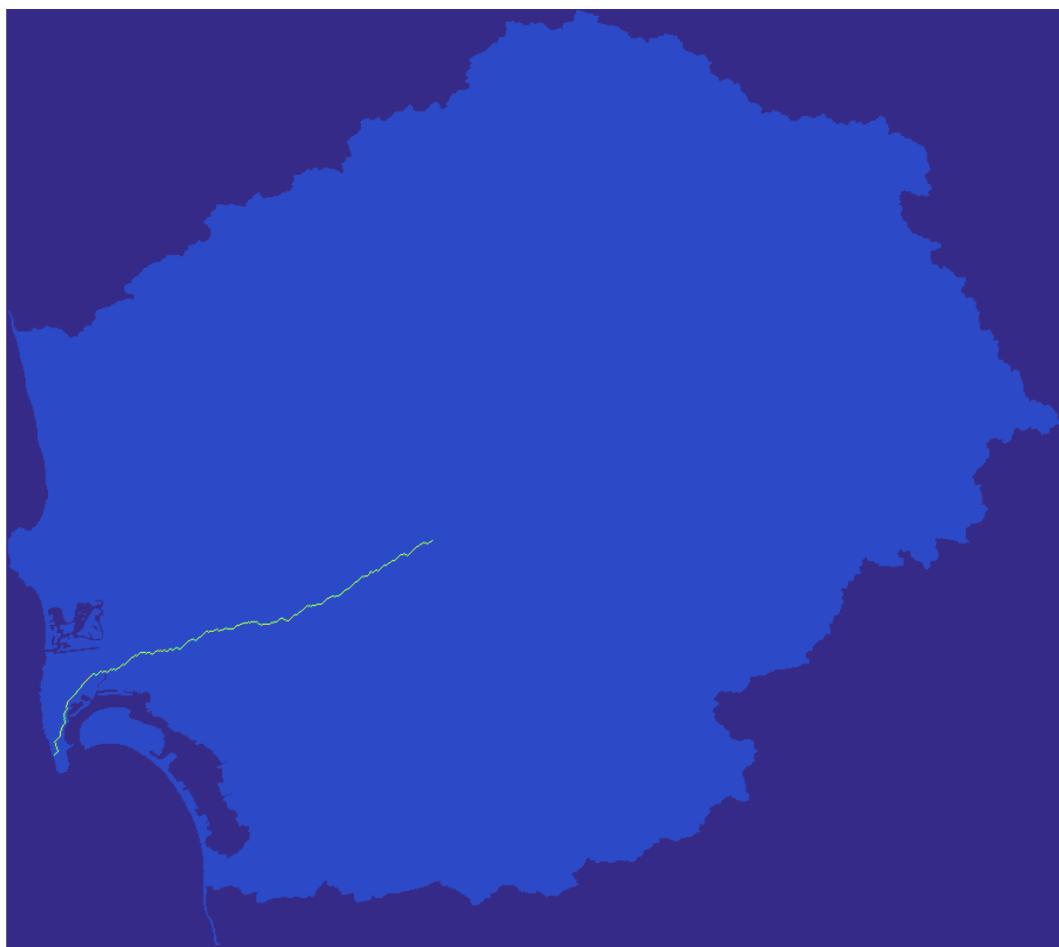


Figure 4.34: San Diego Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview

- Total Area: $6,943.6 \text{ km}^2$
- Maximum Elevation: $1,536.6 \text{ m}$
- Minimum Elevation: 0 m
- Mean Slope: 2.16%
- Standard Deviation of Slope: 6.24%
- Dominant Soil Composition: Hydrologic Soil Group - B: 10 – 20% clay, 50 – 90% sand, 35% rock fragments



Figure 4.35: Santa Ana – San Bernadino Region Overview (Filled in Black)

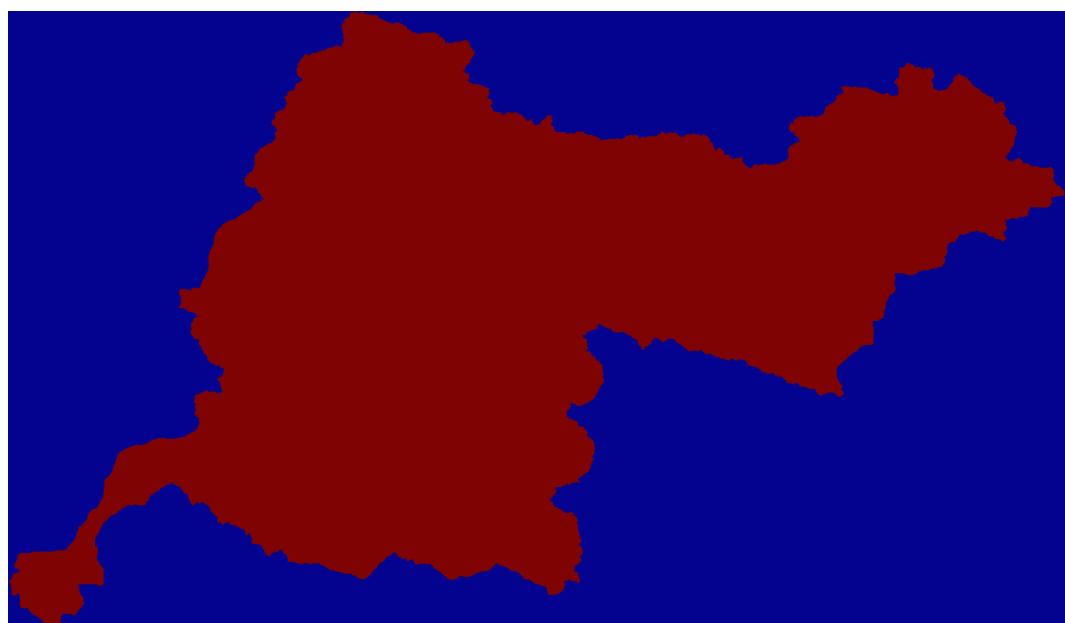


Figure 4.36: Santa Ana – San Bernadino Region Search Domain (Filled in Red)

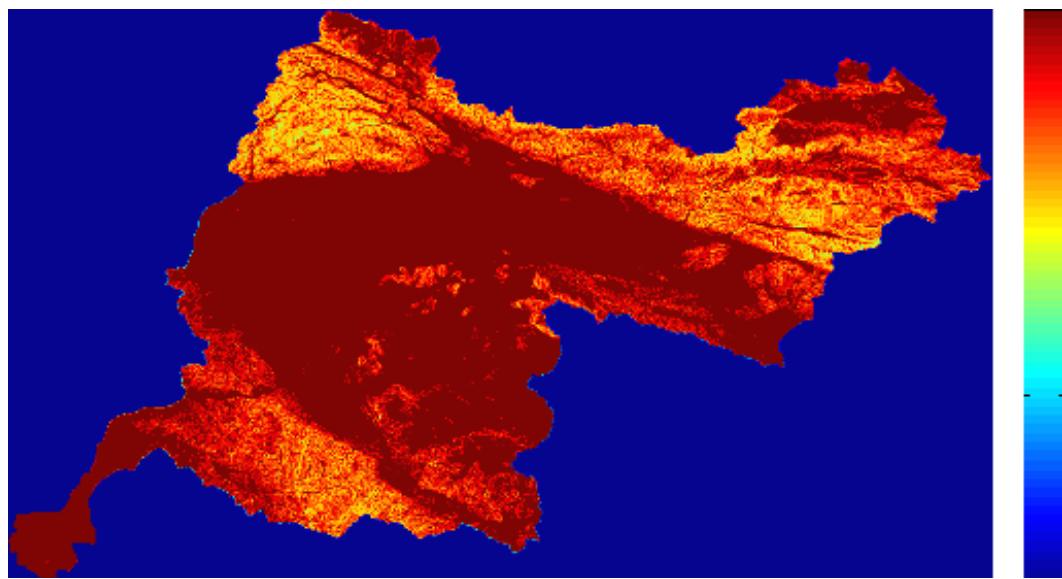


Figure 4.37: Santa Ana – San Bernadino Region Destination Search Inputs: Slope Score (Blue:Low, Red:High)

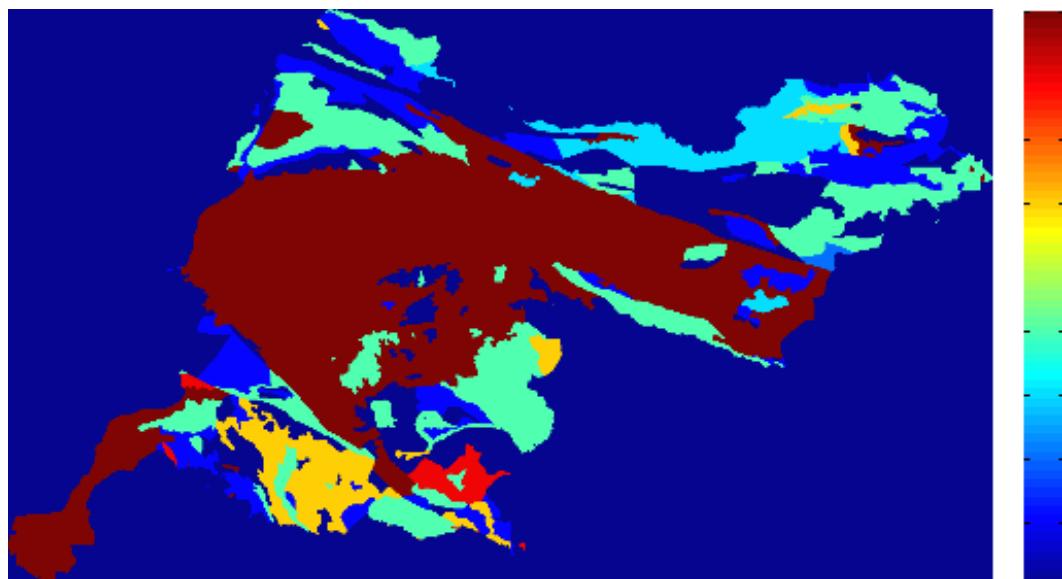


Figure 4.38: Santa Ana – San Bernadino Region Destination Search Inputs: Geology Score (Blue:Low, Red:High)

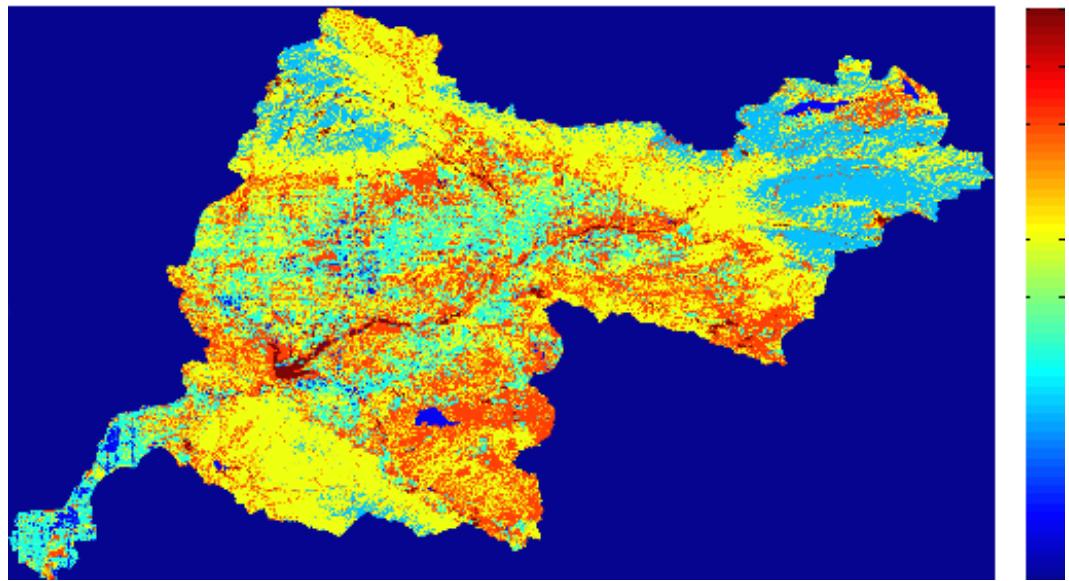


Figure 4.39: Santa Ana – San Bernadino Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High)

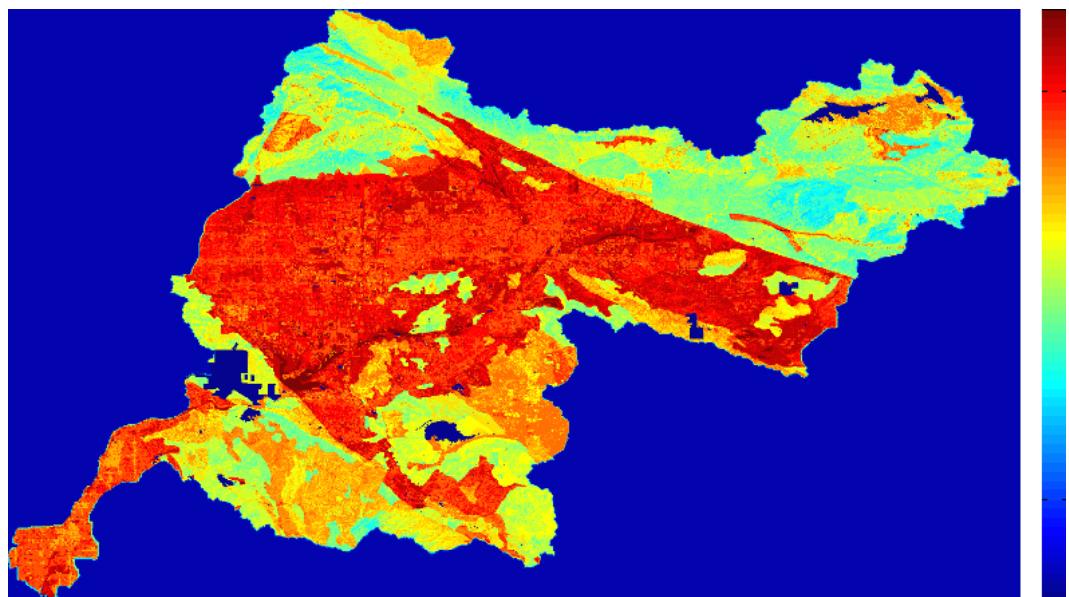


Figure 4.40: Santa Ana – San Bernadino Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)

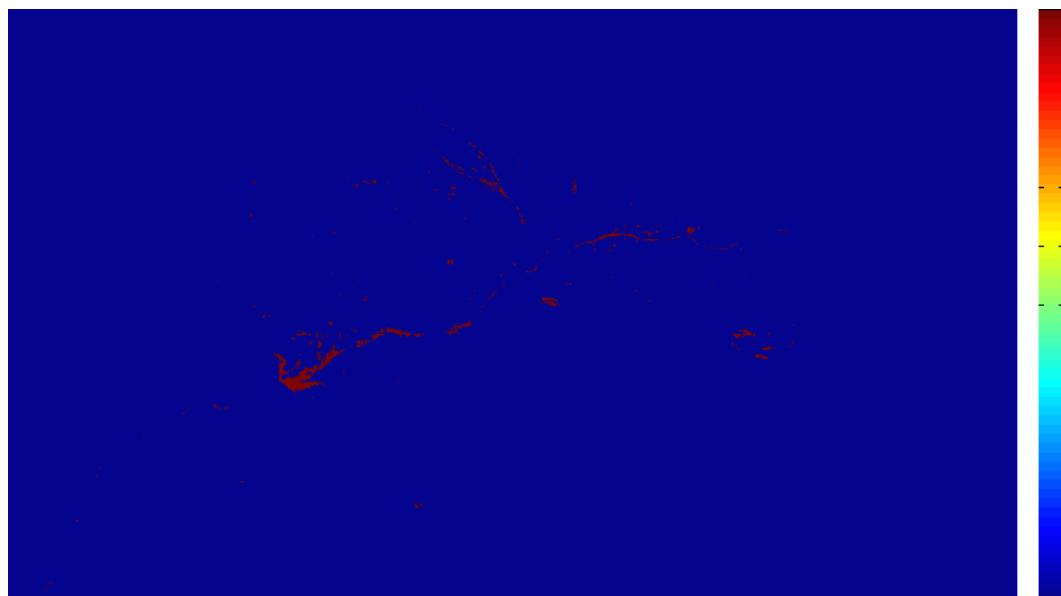


Figure 4.41: Santa Ana – San Bernadino Region Destination Search Outputs: Candidate Regions

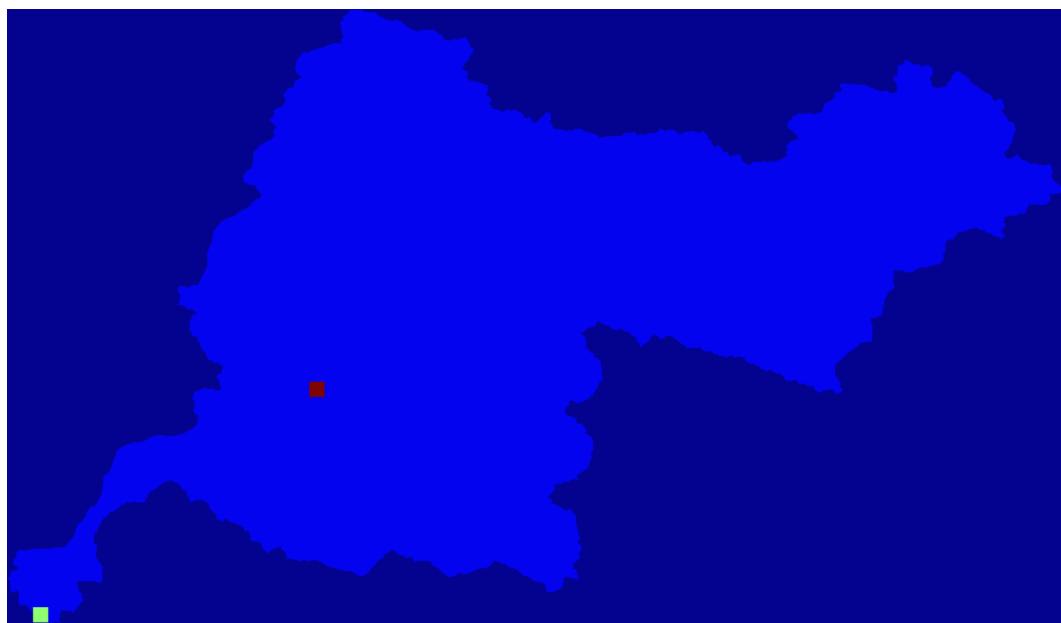


Figure 4.42: Santa Ana – San Bernadino Region Proposed Corridor Endpoints

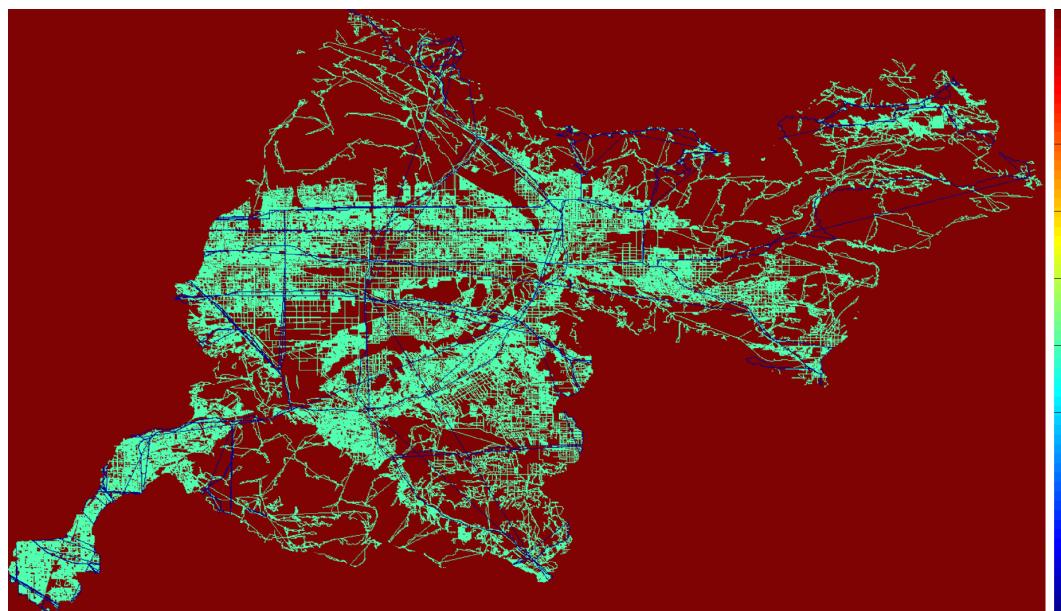


Figure 4.43: Santa Ana – San Bernadino Region Accessibility Based Objective Scores (Blue:Low, Red:High)

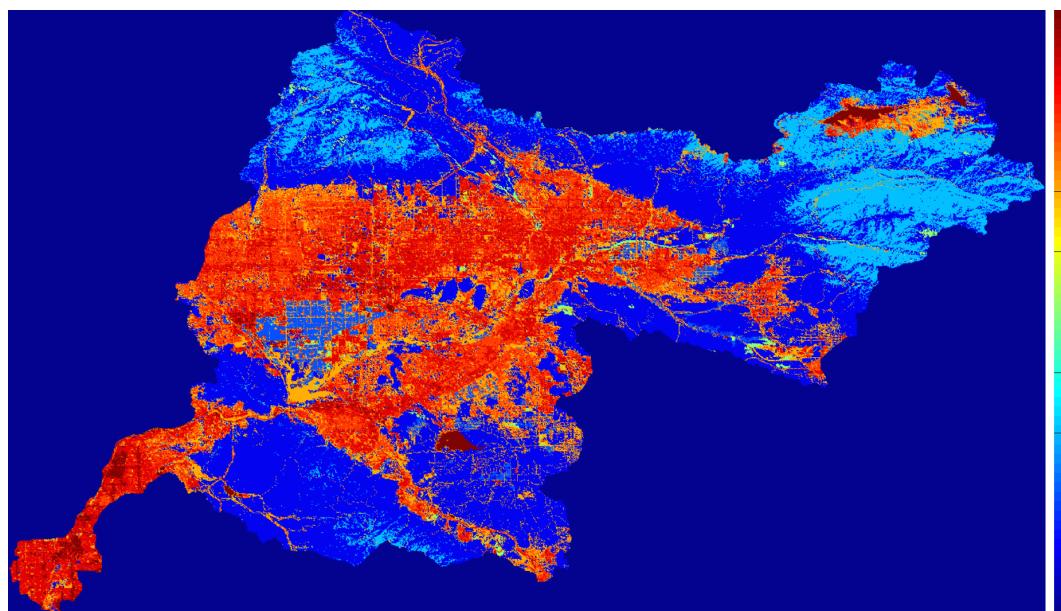


Figure 4.44: Santa Ana – San Bernadino Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)

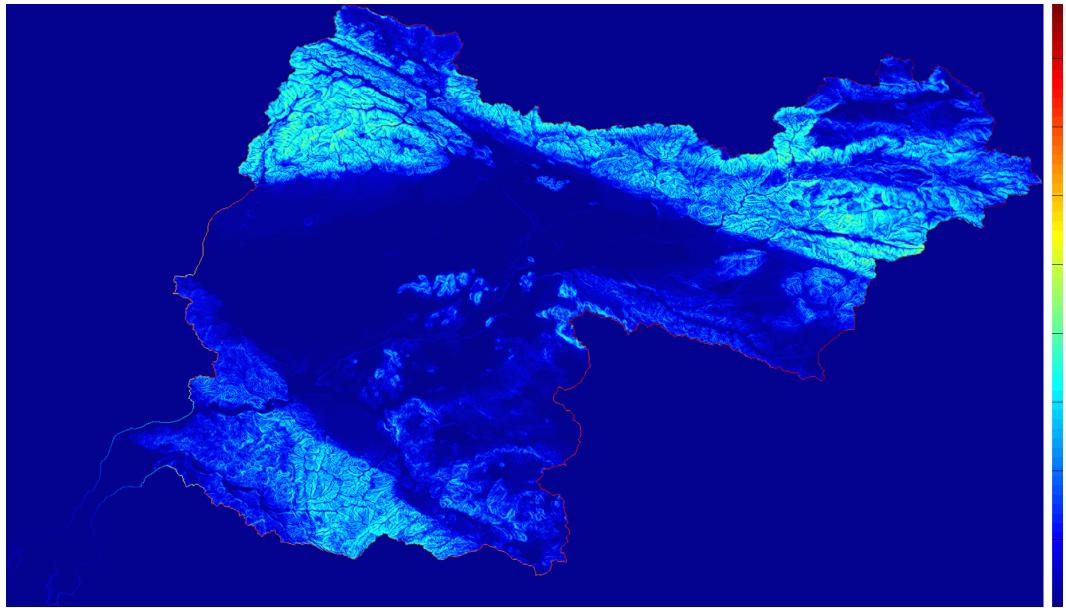


Figure 4.45: Santa Ana – San Bernardino Region Slope Based Objective Scores (Blue:Low, Red:High)

4.5.2 SEARCH DOMAIN

- Grid Dimensions: 1018 *cells* x 1459 *cells*
- Grid Cell Resolution: 100 m x 100 m (1 ha)
- Feasible Grid Cells: 694,365 *cells*

4.5.3 DESTINATION SEARCH INPUTS

4.5.4 DESTINATION SEARCH OUTPUTS

4.5.5 PROPOSED CORRIDOR ENDPOINTS

- Start Location: (435, 1037)
- End Destination:

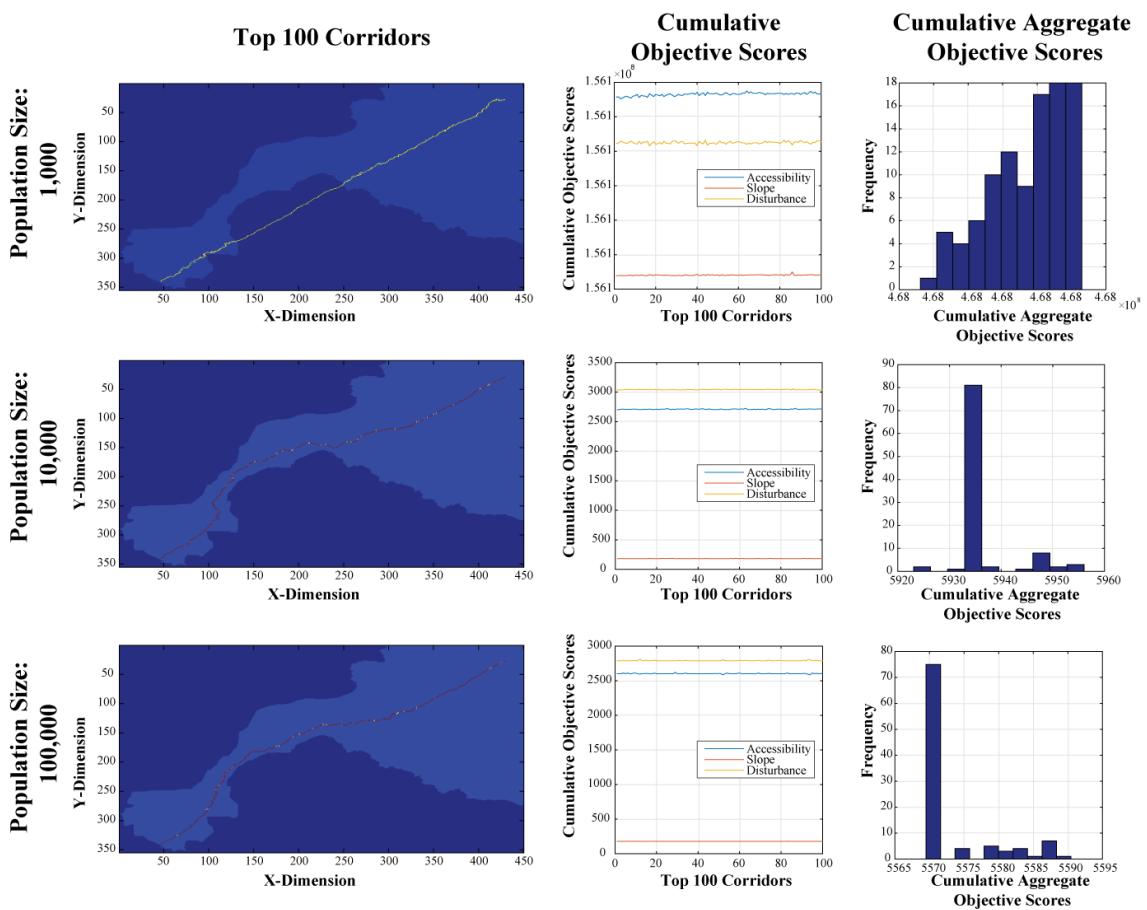


Figure 4.46: Santa Ana – San Bernardino Region Corridor Analysis Results

- Shortest Euclidean Path Distance:

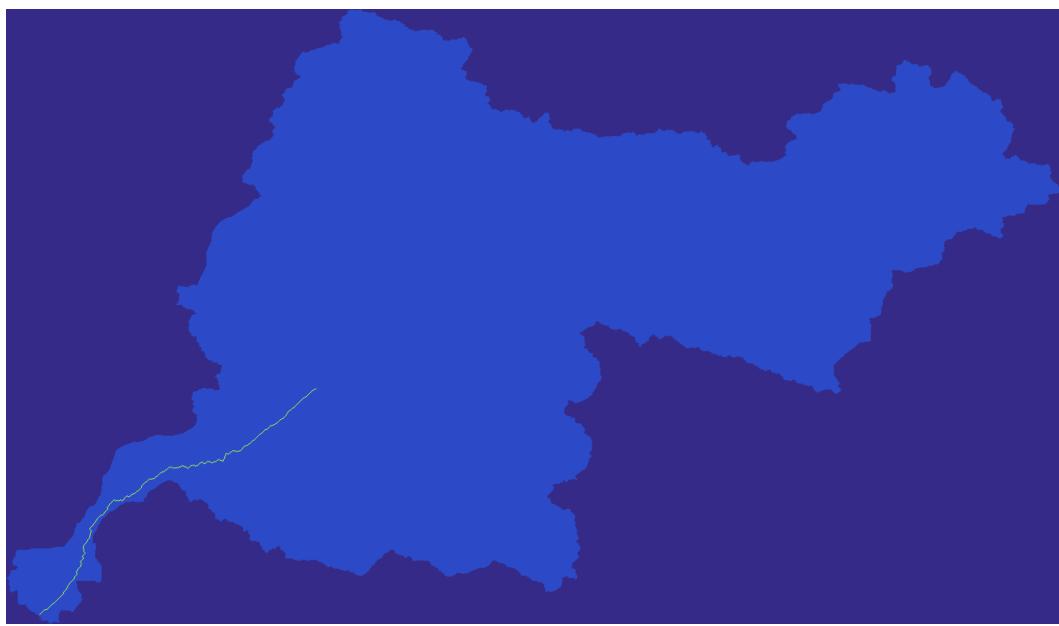


Figure 4.47: Santa Ana – San Bernadino Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview



Figure 4.48: Fresno – Tulare Region Overview (Filled in Black)

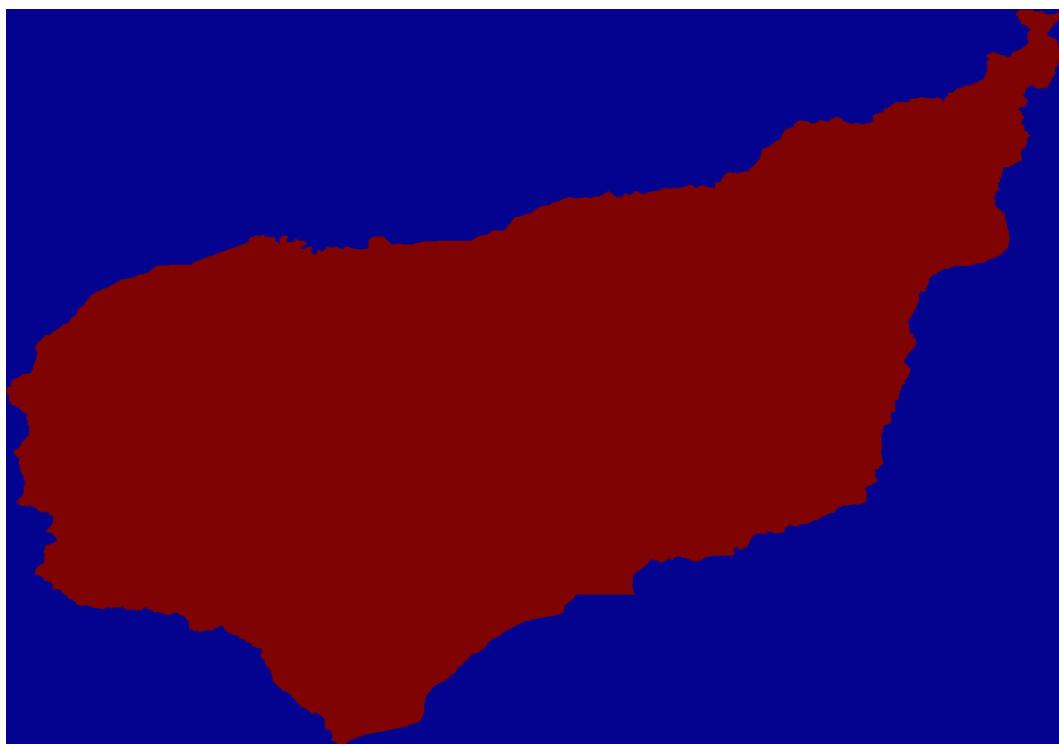


Figure 4.49: Fresno – Tulare Region Search Domain (Filled in Red)

4.5.6 PROPOSED OBJECTIVE LAYERS

4.5.7 PROPOSED CORRIDOR SOLUTIONS

4.5.8 ANTICIPATED DISTRIBUTION OF LIFE CYCLE ENERGY USAGES AND NET WATER SAVINGS

4.6 EVALUATING ALGORITHM RUNTIME PERFORMANCE

4.7 EVALUATING ALGORITHM SOLUTION QUALITY

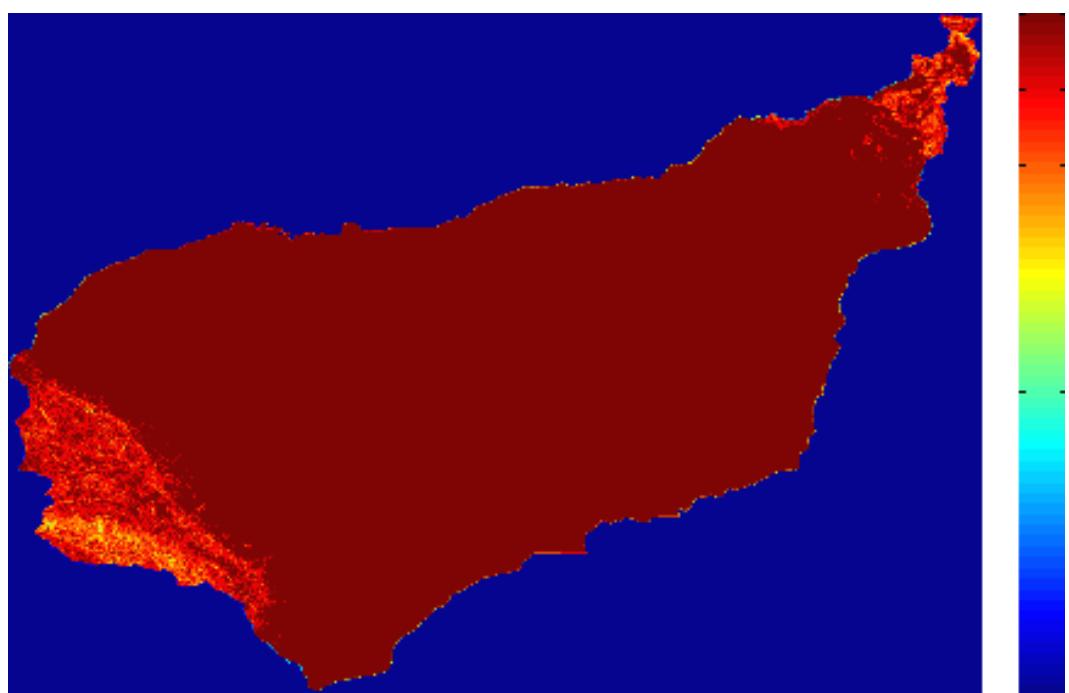


Figure 4.50: Fresno – Tulare Region Destination Search Inputs: Slope Score (Blue:Low, Red:High)



Figure 4.51: Fresno – Tulare Region Destination Search Inputs: Geology Score (Blue:Low, Red:High)

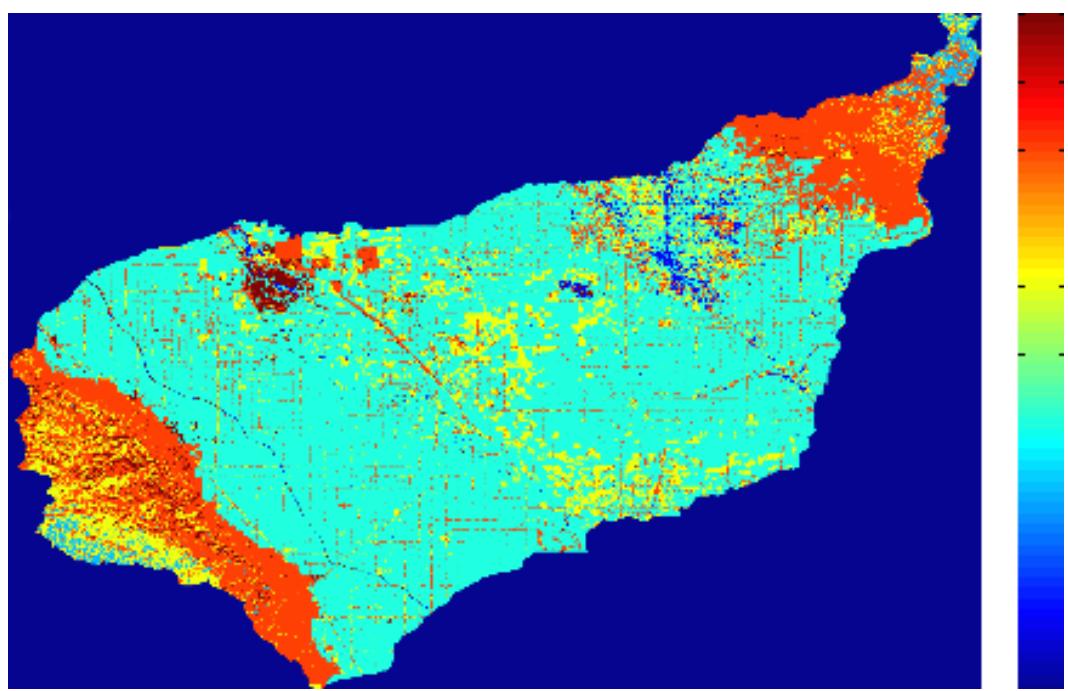


Figure 4.52: Fresno – Tulare Region Destination Search Inputs: Landuse Score (Blue:Low, Red:High)

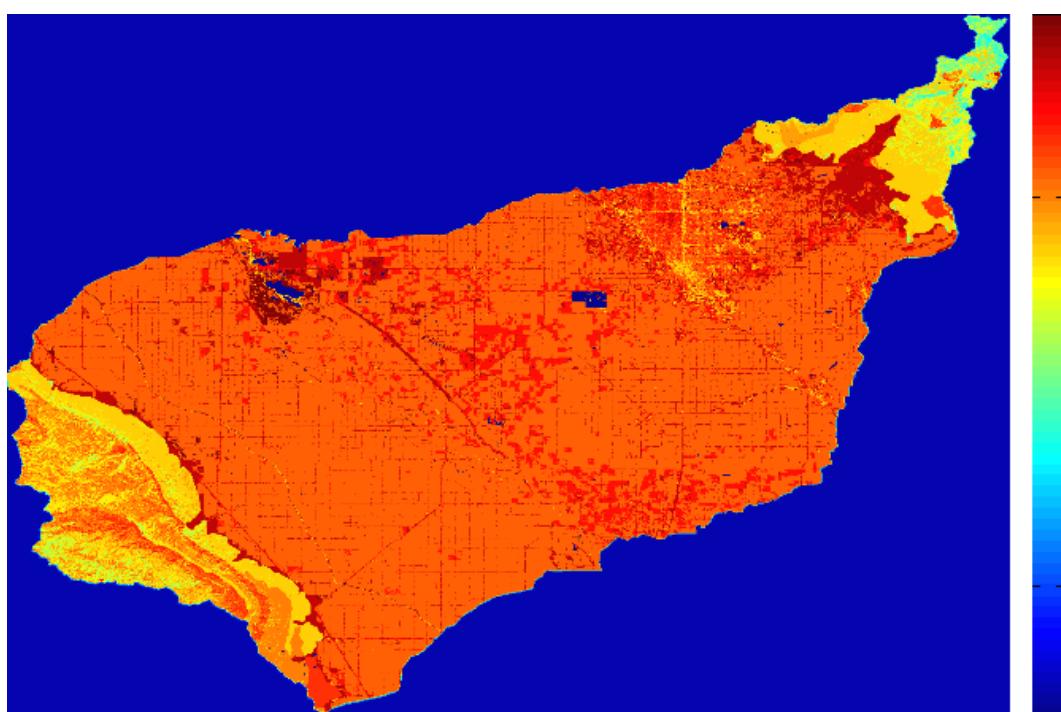


Figure 4.53: Fresno – Tulare Region Destination Search Outputs: Composite Scores (Blue:Low, Red:High)

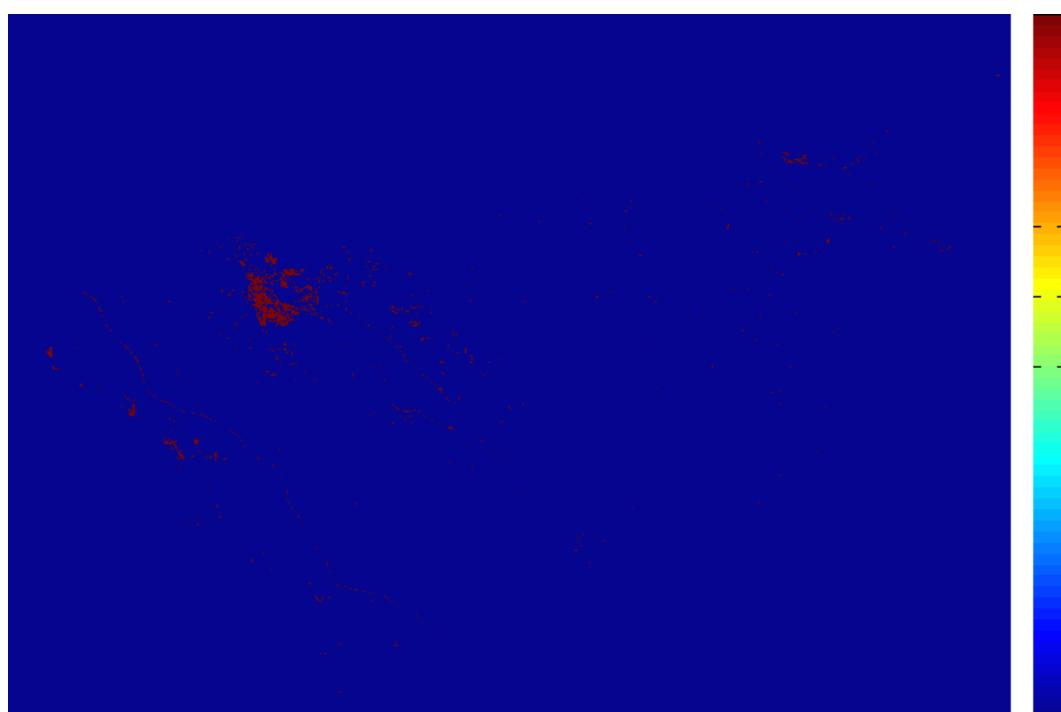


Figure 4.54: Fresno – Tulare Region Destination Search Outputs: Candidate Regions



Figure 4.55: Fresno – Tulare Region Proposed Corridor Endpoints

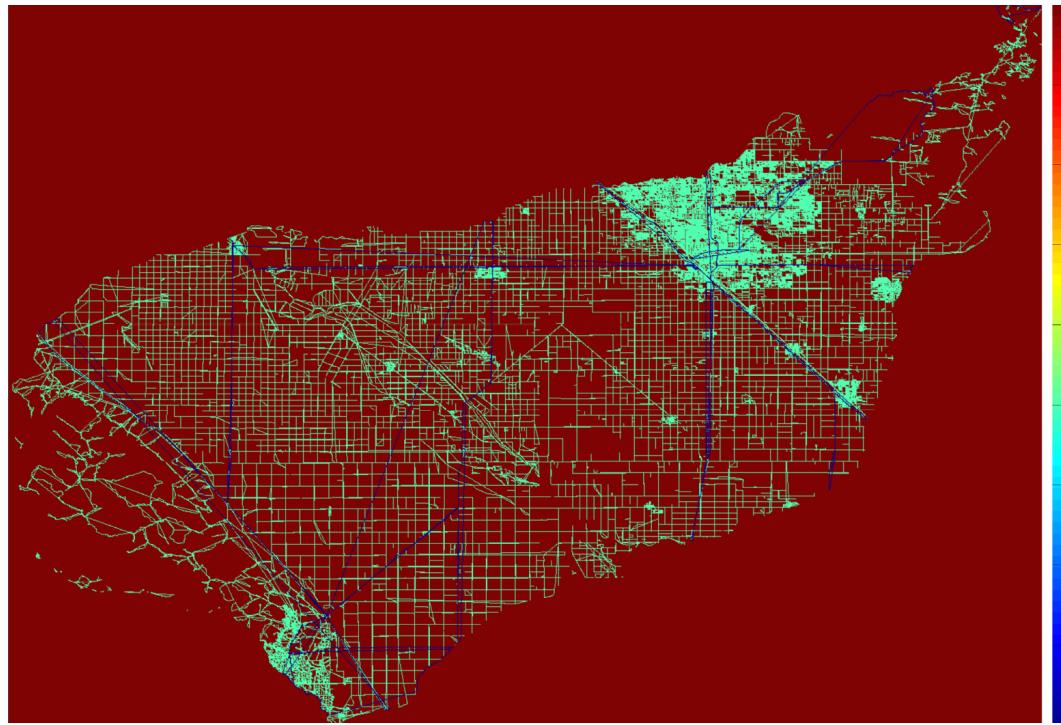


Figure 4.56: Fresno – Tulare Region Accessibility Based Objective Scores (Blue:Low, Red:High)

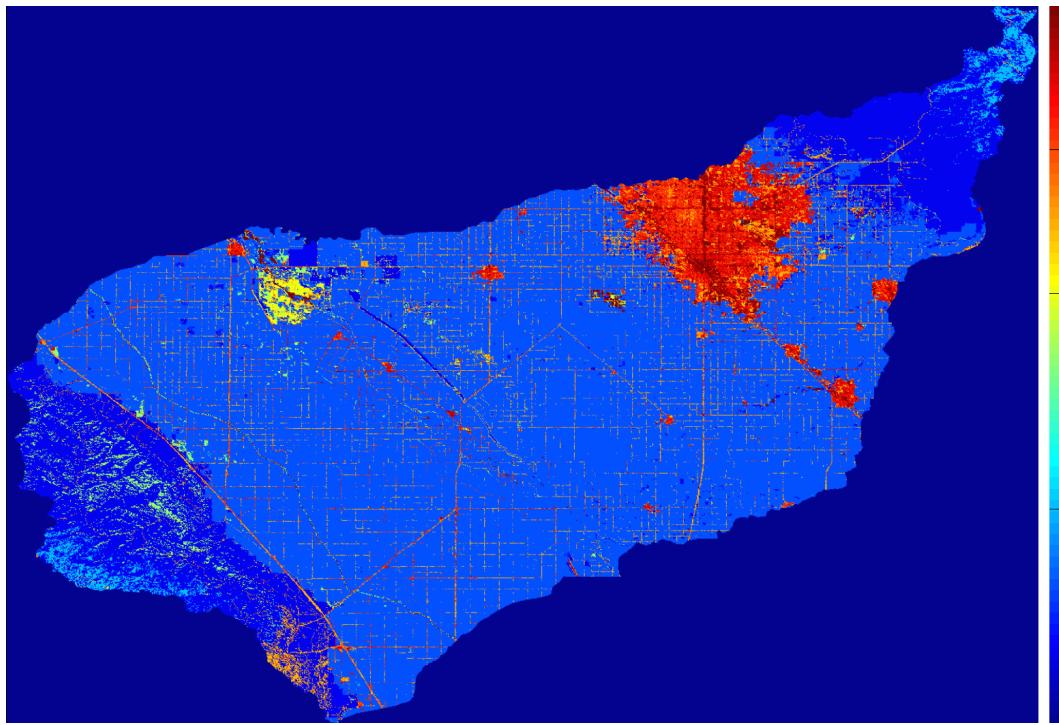


Figure 4.57: Fresno – Tulare Region Land Use Disturbance Based Objective Scores (Blue:Low, Red:High)

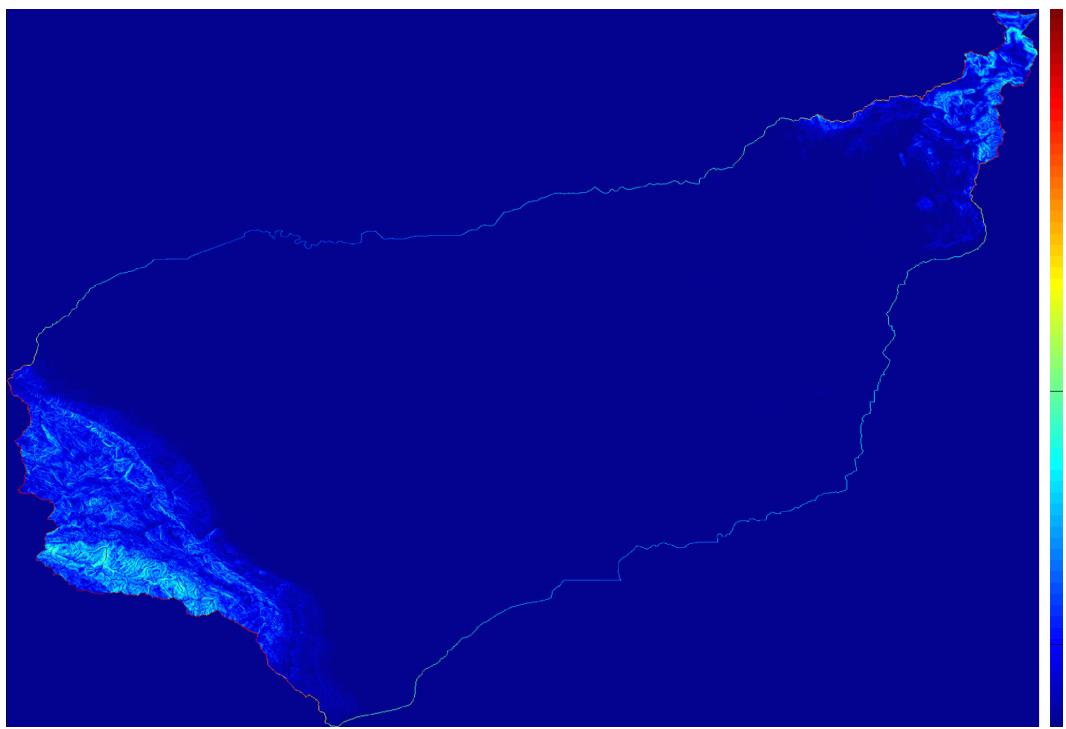


Figure 4.58: Fresno – Tulare Region Slope Based Objective Scores (Blue:Low, Red:High)

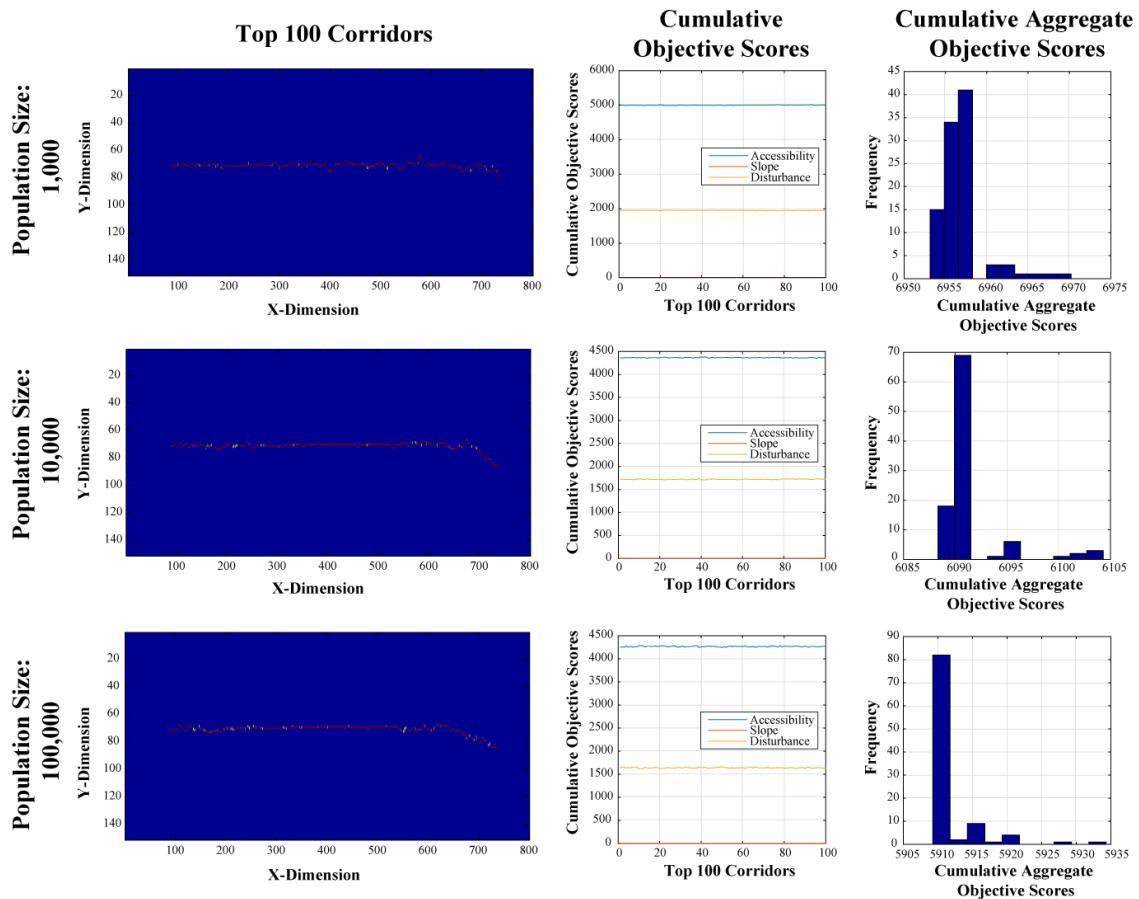


Figure 4.59: Fresno Region Corridor Analysis Results



Figure 4.60: Fresno Region Top 100 Corridors (Pop Size: 100,000) Basin Wide Overview

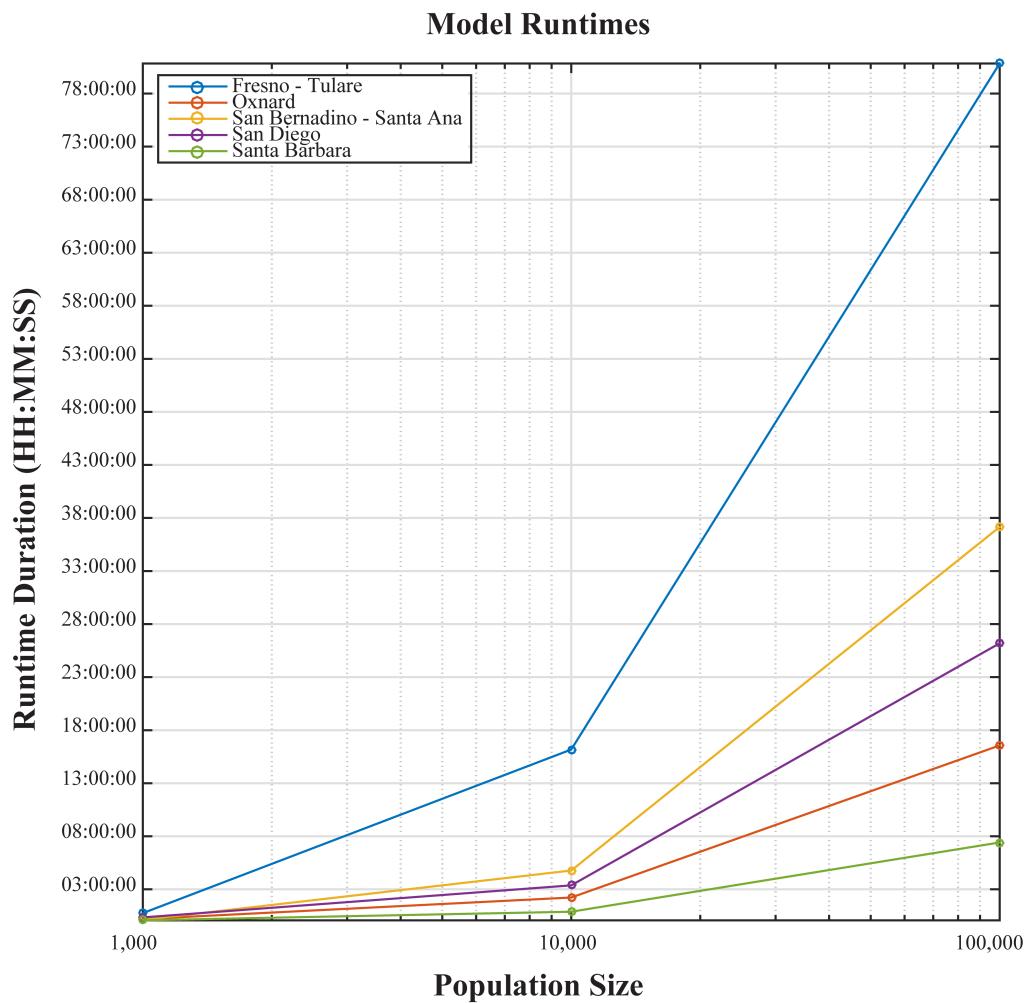


Figure 4.61: Algorithm Runtime Performance for Each of the Five Case Study Regions for Three Population Sizes

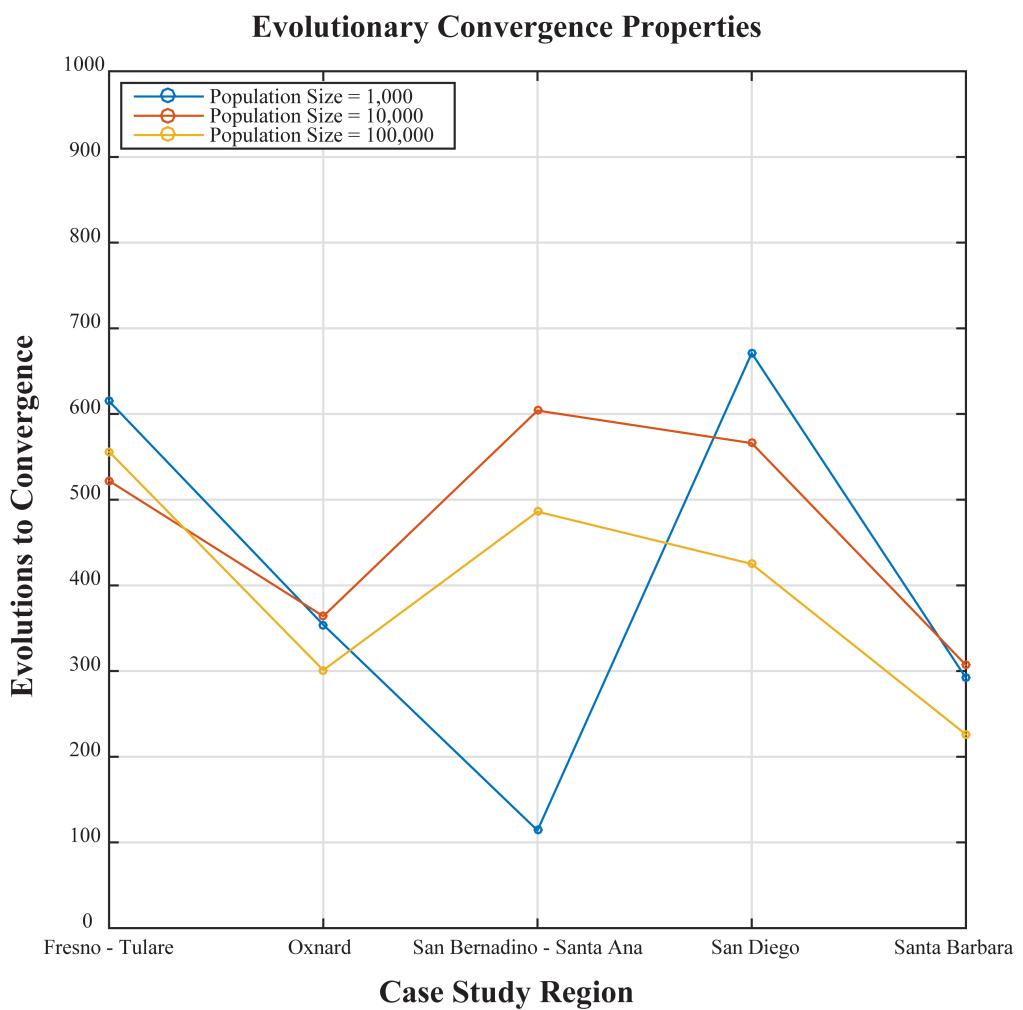


Figure 4.62: Algorithm Convergence Rates for Each of the Five Case Study Regions for Three Population Sizes

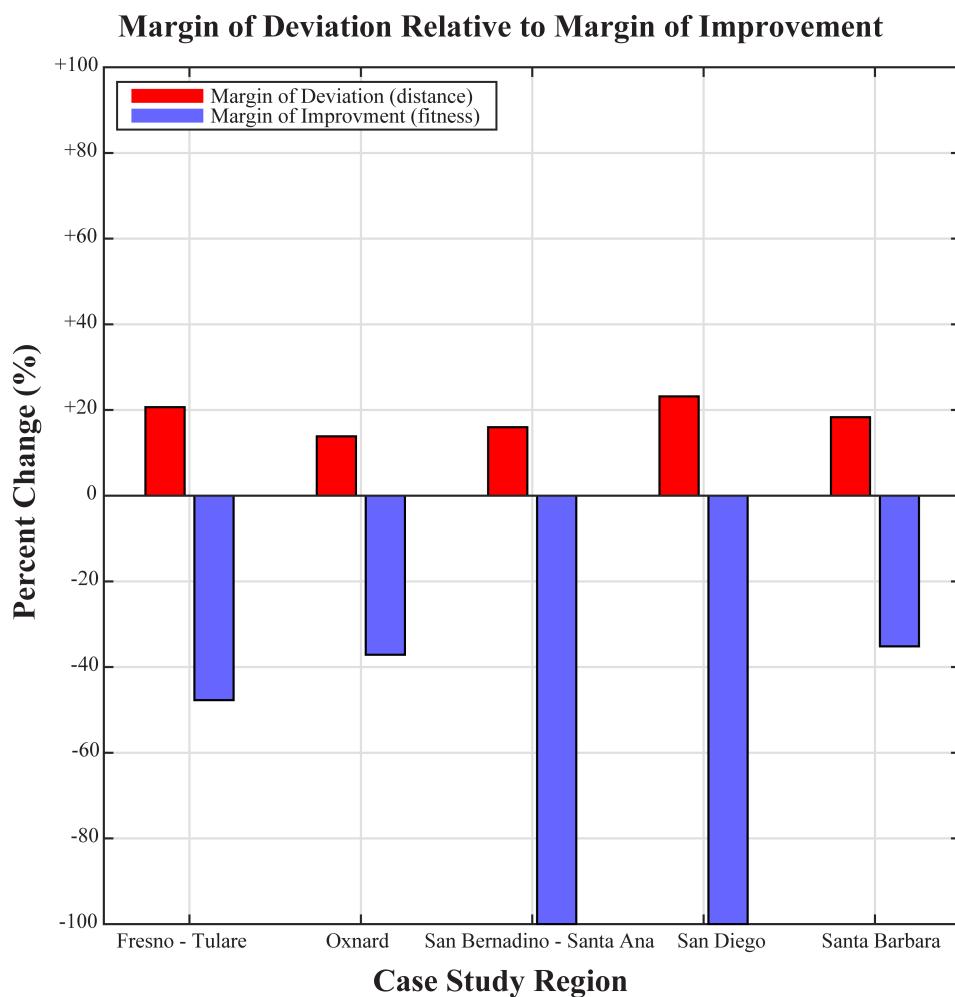


Figure 4.63: Comparison of the Along Path Distance and Cumulative Objective Scores between the Solution Corridors and the Euclidean Shortest Corridors

References

- [1] Aissi, H., Chakhar, S., & Mousseau, V. (2012). GIS-based multicriteria evaluation approach for corridor siting. *Environment and Planning B: Planning and Design*, 39(2), 287–307.
- [2] Bennett, D. A., Xiao, N., & Armstrong, M. P. (2004). Exploring the geographic consequences of public policies using evolutionary algorithms. *Annals of the Association of American Geographers*, 94(4), 827–847.
- [3] Bolstad, P. (2005). *GIS Fundamentals: A First Text on Geographic Information Systems*. Eider Press.
- [4] Chakhar, S. & Martel, J.-M. (2003). Enhancing geographical information systems capabilities with multi-criteria evaluation functions. *Journal of Geographic Information and Decision Analysis*, 7(2), 47–71.
- [5] Church, R. L. (2002). Geographical information systems and location science. *Computers & Operations Research*, 29(6), 541–562.
- [6] Church, R. L., Loban, S. R., & Lombard, K. (1992). An interface for exploring spatial alternatives for a corridor location problem. *Computers & Geosciences*, 18(8), 1095–1105.
- [7] Coello, C., Lamont, G., & Veldhuisen, D. V. (2007). *Evolutionary Algorithms for Solving Multi-Objective Problems*. New York, NY: Springer, second edi edition.
- [8] Coello, C. A. C., Aguirre, A. H., & Zitzler, E. (2001). Evolutionary multi-criterion optimization.
- [9] Collins, M. G., Steiner, F. R., & Rushman, M. J. (2001). Land-use suitability analysis in the United States: Historical development and promising technological achievements. *Environmental Management*, 28(5), 611–621.
- [10] Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. New York, New York, USA: Wiley.
- [11] Fonseca, C. M. & Fleming, P. J. (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary computation*, 3(1), 1–16.

- [12] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, volume Addison-Wesley of *Artificial Intelligence*. Addison-Wesley.
- [13] Goldberg, D. E. & Holland, J. H. (1988). Genetic Algorithms and Machine Learning. *Machine Learning*, 3(2), 95–99.
- [14] Goodchild, M. F. (1977). An evaluation of lattice solutions to the problem of corridor location. *Environment and Planning A*, 9(7), 727–738.
- [15] Hallam, C., Harrison, K. J., & Ward, J. A. (2001). A multiobjective optimal path algorithm. *Digital Signal Processing*, 11(2), 133–143.
- [16] Hopkins, L. D. (1977). Methods for Generating Land Suitability Maps: A Comparative Evaluation. *Journal of the American Institute of Planners*, 43(4), 386–400.
- [17] Huber, D. L. & Church, R. L. (1985). Transmission corridor location modeling. *Journal of Transportation Engineering*, 111(2), 114–130.
- [18] Jankowski, P. (1995). Integrating geographical information systems and multiple criteria decision-making methods. *International journal of geographical information systems*, 9(3), 251–273.
- [19] Lombard, K. & Church, R. L. (1993). The gateway shortest path problem: generating alternative routes for a corridor location problem. *Geographical systems*, 1(1), 25–45.
- [20] Malczewski, J. (2004). GIS-based land-use suitability analysis: a critical overview. *Progress in Planning*, 62(1), 3–65.
- [21] Malczewski, J. (2006). GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20(7), 703–726.
- [22] Mooney, P. & Winstanley, A. (2006). An evolutionary algorithm for multicriteria path optimization problems. *International Journal of Geographical Information Science*, 20(4), 401–423.
- [23] Mousseau, V., Aissi, H., & Chakhar, S. (2010). A Three-Phase Approach and a Fast Algorithm to Compute Efficient Corridors within GIS Framework. *CER*, 10(7), 1–18.
- [24] Neema, M. & Ohgai, a. (2010). Multi-objective location modeling of urban parks and open spaces: Continuous optimization. *Computers, Environment and Urban Systems*, 34(5), 359–376.
- [25] Roberts, S. a., Hall, G. B., & Calamai, P. H. (2010). Evolutionary Multi-objective Optimization for landscape system design. *Journal of Geographical Systems*, 13(3), 299–326.

- [26] Scaparra, M. P., Church, R. L., & Medrano, F. A. (2014). Corridor location: the multi-gateway shortest path model. *Journal of Geographical Systems*, 16(3), 287–309.
- [27] Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). *Hydrologic Unit Maps: US Geological Survey Water Supply Paper 2294*. USGS.
- [28] Stefanakis, E. & Kavouras, M. (1995). On the determination of the optimum path in space. In *Spatial Information Theory A Theoretical Basis for GIS* (pp. 241–257). Springer.
- [29] Tsai, C.-C., Huang, H.-C., & Chan, C.-K. (2011). Parallel elite genetic algorithm and its application to global path planning for autonomous robot navigation. *Industrial Electronics, IEEE Transactions on*, 58(10), 4813–4821.
- [30] Zhang, X. & Armstrong, M. P. (2008). Genetic algorithms and the corridor location problem: multiple objectives and alternative solutions. *Environment and Planning B: Planning and Design*, 35(1), 148–168.
- [31] Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z., Suganthan, P. N., & Zhang, Q. (2011). Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, 1(1), 32–49.
- [32] Zhou, J. & Civco, D. (1996). Using genetic learning neural networks for spatial decision making in GIS. *Photogrammetric Engineering and Remote Sensing*, 62(11), 1287–1295.
- [33] Zitzler, E. & Thiele, L. (1998). Multiobjective optimization using evolutionary algorithms—a comparative case study. In *Parallel problem solving from nature—PPSN V* (pp. 292–301). Springer.



THIS THESIS WAS TYPESET using L^AT_EX,
originally developed by Leslie Lamport
and based on Donald Knuth's T_EX. The
body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist
typeface.