

University of Redlands

InSPIRe @ Redlands

---

MS GIS Program Major Individual Projects

Theses, Dissertations, and Honors Projects

---

10-2013

## A Geographic Information System for Invasive Species: Sahara Mustard Weed

Violet M. Cullors  
*University of Redlands*

Follow this and additional works at: [https://inspire.redlands.edu/gis\\_gradproj](https://inspire.redlands.edu/gis_gradproj)



Part of the [Geographic Information Sciences Commons](#)

---

### Recommended Citation

Cullors, V. M. (2013). *A Geographic Information System for Invasive Species: Sahara Mustard Weed* (Master's thesis, University of Redlands). Retrieved from [https://inspire.redlands.edu/gis\\_gradproj/205](https://inspire.redlands.edu/gis_gradproj/205)



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material may be protected by copyright law (Title 17 U.S. Code).

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Honors Projects at InSPIRe @ Redlands. It has been accepted for inclusion in MS GIS Program Major Individual Projects by an authorized administrator of InSPIRe @ Redlands. For more information, please contact [inspire@redlands.edu](mailto:inspire@redlands.edu).

University of Redlands

**A Geographic Information System for Invasive Species:  
Sahara Mustard Weed**

A Major Individual Project submitted in partial satisfaction of the requirements  
for the degree of Master of Science in Geographic Information Systems

by

Violet Cullors

Rujin Ma, Ph.D., Committee Chair

Mark Kumler, Ph.D.

October 2013

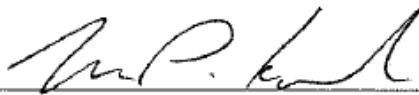
A Geographic Information System for Invasive Species: Sahara Mustard Weed

Copyright © 2013

by

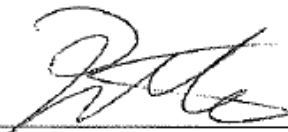
Violet Cullors

The report of Violet Cullors is approved.



---

Mark Kumler, Ph.D.



---

Rujin Ma, Ph.D., Committee Chair

October 2013



## Acknowledgements

Foremost, I would like to thank my parents for inspiring me to not ever quit. You have always provided unyielding love, commitment, strength, and hard work to our home, our family, and our country. I achieved because you believed in me.

It is truly an honor to be selected by NGA to attend the GIS master's program; I extend my deepest gratitude for their generous support. I send a big warm thank you to the friends I made in Cohort 22 and the University of Redlands GIS Master's staff for your constant support and camaraderie.

Throughout my life I have had mentors and friends who guided me forward in my life and career. They helped to change my thinking, and as a result I changed my world. In my late forties I didn't think I needed any more mentoring, but Lynda Chamowitz proved me wrong. She exemplifies courage, strength and tenacity. You work tirelessly doing what you believe in, sometimes in difficult circumstances. With 30 years of Government service and as my immediate supervisor at NGA, you provided me unparalleled wisdom, support and friendship. You pushed me out of my comfort zone to perform in ways I didn't think I could. You taught me how to be more effective. You sent me to writing school, despite my reluctance. Because of your determination, I was prepared to start and finish this master's program. Thank you, Lynda, for helping me reach further and higher than I even dreamed possible, I dedicate this work to you.



## **Abstract**

A Geographical Information System for Invasive Species: Sahara Mustard Weed

by

Violet Cullors

Invasive plant species are aggressively spreading and threatening the Joshua Tree National Park (JOTR) ecosystem. Uncontrolled invasives will crowd out native plants and disrupt the natural habitat for desert animal species. JOTR needed a geographic information system (GIS) that can provide support for developing weed control plans against the Sahara mustard threat. This project addressed this need by developing a geodatabase for analysis, compiling required GIS feature layers, developing a mustard weed data model and a predictive spread model to aid in tracking the invasive weed. The data model identifies the essential data to collect for assessing and monitoring mustard weed observations. The compiled GIS feature layers consists of human activity factors (road network, trail, disturbed areas) and land factors (soil type, elevation, slope, vegetation cover, etc.). Human activity is a strong predictor of weed spread and these feature classes are the main element in one tool, the Predict Weed Spread Model. Land assessment analysis helps identify JOTR areas that are potentially high risk to mustard weed infestation. Results showed that a fundamental understanding of the Sahara mustard dynamics is required to model weed habitats and to predict weed spread that contributed to its existence and spread.





# Table of Contents

<b>Chapter 1 – Introduction .....</b>	<b>1</b>
1.1 Client.....	2
1.2 Problem Statement .....	2
1.3 Proposed Solution .....	3
1.3.1 Goals and Objectives .....	3
1.3.2 Scope.....	3
1.3.3 Methods.....	3
1.4 Audience .....	4
1.5 Overview of the Rest of This Report .....	4
<b>Chapter 2 – Background and Literature Review .....</b>	<b>5</b>
2.1 Invasive Sahara Mustard Weed .....	5
2.2 Predictive Habitat Distribution Models .....	6
2.3 Remote Sense Datasets for Habitat Suitability Models .....	8
2.4 Summary .....	8
<b>Chapter 3 – Systems Analysis and Design.....</b>	<b>9</b>
3.1 Problem Statement .....	9
3.2 Requirements Analysis .....	9
3.2.1 Data Requirements.....	9
3.2.2 System Functionality Requirements .....	10
3.2.3 Non-functional Requirements.....	11
3.3 System Design .....	11
3.4 Project Plan .....	12
3.5 Summary .....	13
<b>Chapter 4 – Database Design.....</b>	<b>15</b>
4.1 Conceptual Model.....	15
4.2 Logical Data Model .....	16
4.2.1 Sahara Mustard Weed Feature Class .....	16
4.2.2 JOTR Land Layers .....	17
4.2.3 Byway Layers .....	18
4.2.4 Other Useful Layers .....	19
4.3 Data Source.....	20
4.4 Data Collection Methods .....	21
4.5 Data Scrubbing and Loading .....	21
4.6 Summary .....	22
<b>Chapter 5 – Implementation.....</b>	<b>23</b>
5.1 Geodatabase .....	23
5.2 Land Assessment using the Geodatabase .....	26
5.3 Predictive Weed Spread Model .....	27
5.4 Summary .....	35
<b>Chapter 6 – Results and Analysis.....</b>	<b>37</b>
6.1 Land Assessment Analysis Results.....	37

6.2	Results of the Predictive Spread Model.....	40
6.3	Summary.....	44
<b>Chapter 7 – Conclusions and Future Work.....</b>		<b>45</b>
7.1	Summary.....	45
7.2	Future Work.....	45
<b>Works Cited .....</b>		<b>67</b>
<b>Appendix A. Python String Function.....</b>		<b>49</b>
<b>Appendix B. Rescale Tool Code .....</b>		<b>50</b>
<b>Appendix C. Region Growing Code.....</b>		<b>51</b>

## Table of Figures

Figure 1-1: Map of Joshua Tree National Park, California .....	1
Figure 1-2: Cholla Cactus Garden (NPS U.S. Department of the Interior, 2012) .....	2
Figure 3-1: System Design.....	12
Figure 3-2: Project Gantt Chart.....	13
Figure 4-1: Conceptual Model of the Sahara Mustard Weed .....	15
Figure 4-2: Close-up View of Disturbed Areas near JOTR's Park Blvd. ....	18
Figure 4-3: Elevation Layer over JOTR .....	18
Figure 4-4: JOTR Roads Layer.....	19
Figure 4-5: Overview of USGS Quadrangle Tile Names .....	20
Figure 5-1: Selection Tool and SQL Used to Extract Disturbed Areas Features .....	24
Figure 5-2: Disturbed Areas Types.....	25
Figure 5-3: Definition Query of the SQL Statement for Display .....	27
Figure 5-4: Data Flow of First Phase of Predictive Spread Implementation .....	28
Figure 5-5: Data Flow for Second Phase of the Predictive Spread Implementation .....	29
Figure 5-6: Scale Values from Threshold to Maximum Scale Number .....	30
Figure 5-7: Rescale Elevation Tool Interface .....	30
Figure 5-8: Roads Euclidean Distance.....	31
Figure 5-9: Sahara Mustard Weed Euclidean Distances.....	32
Figure 5-10: Example of a Resulting Impedance Surface Layer .....	33
Figure 5-11: Unconnected Cells (i.e. dash lines) within Range of a Weed Point.....	34
Figure 5-12: Revised Data Flow for Second Phase of Predictive Spread.....	35
Figure 6-1: Interface for Feature Layers for Land Assessment Analysis .....	37
Figure 6-2: User Interface of SelectSoils Tool .....	38
Figure 6-3: JOTR Overview with Selected Soil and Elevation Classes .....	39
Figure 6-4: Identified Potential High Risk Areas for Sahara Mustard Invasion.....	39
Figure 6-5: Only Nine Mustard Weed Points in High Risk Areas .....	40
Figure 6-6: User Interface for Script SelectByways.....	41
Figure 6-7: Rescaled Elevation.....	41
Figure 6-8: Predictive Spread Interface .....	42
Figure 6-9: 100 Meter per Year (1, 3, 5) Prediction Spread Results .....	42
Figure 6-10: Elevation Class Change Impacts the Spread Rate .....	43
Figure 6-11: Example of the Spread Going Beyond the 5 Year Maximum Spread Rate .....	44



## List of Tables

Table 3-1: Function and Non-Functional Requirements .....	10
Table 4-1: Sahara Mustard Feature Class Attributes .....	16
Table 4-2: SSURGO Prefixes Identify the Referred Spatial Entity.....	21
Table 5-1: Feature Classes in the Project Database .....	26



## **List of Acronyms and Definitions**

AIS	Aerial Information Systems
BLM	Bureau of Land Management
FS	Forest Service
GARP	Genetic Algorithm Rule set-Production
GIS	Geographical Information System
JOTR	Joshua Tree National Park
MBCA	Morongo Basin Conversation Association
NGA	National Geospatial-Intelligence Agency
NPS	National Park Service
NRCS	Natural Resources Conservation Service
RSAC	Remote Sensing Applications Center
SSURGO	Soil Survey Geographic
USDA	United States Department of Agriculture
USGS	United States Geological Survey
WERC	Western Ecological Research Center

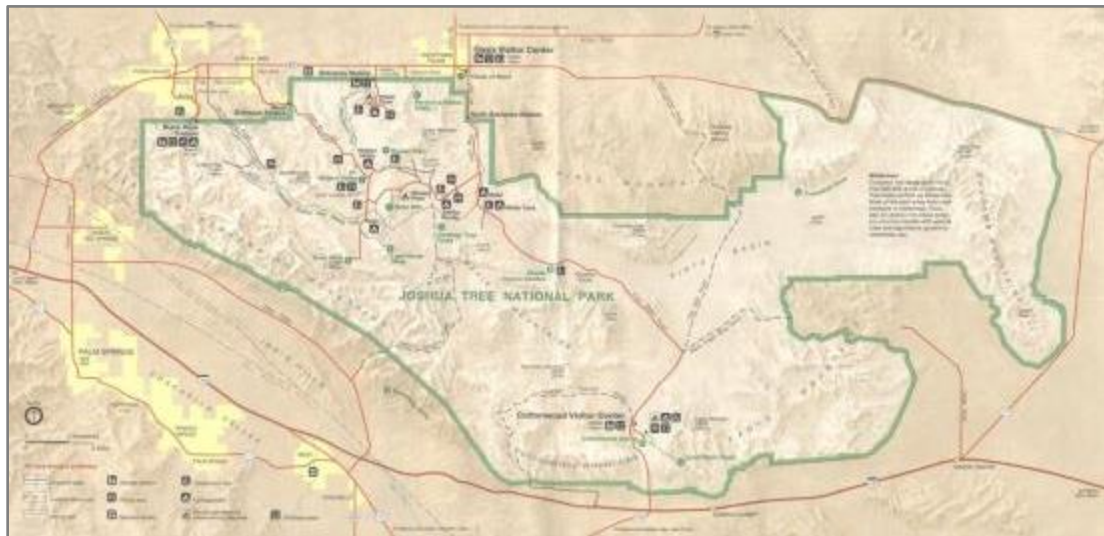




# Chapter 1 – Introduction

The Sahara mustard (*Brassica Tournefortii*) weed is an invasive plant that is decimating native plants in the U.S. southwest region by using available water and nutrients early in the season. Joshua Tree National Park (JOTR) Natural Resource Management team has been battling the Sahara mustard weed’s expansion across the park for years and needed to develop an efficient control strategy for this alien plant infestation. Without control methods in place the Sahara mustard weed can ultimately dominate a desert landscape, displaces native plants, destroy endangered species habitats, and become a potential fire hazard if it accumulates high density populations (USGS Western Ecological Research Center, 2009). Weed factsheets from the University of Nevada (Johnson, 2003) and New Mexico State University (Renz, 2006) state that there are two efficient methods to remove these invasive plants: herbicide and hand pulling. But both methods may produce negative effects to nearby habitats if not carefully planned. JOTR is home to a variety of plants and animals that survive by having an intricate relationship with the desert landscape.

A geographic information system (GIS) can provide decision support information to help JOTR develop a control management plan and to prioritize eradication efforts. GIS would help identify JOTR areas (Figure 1-1) susceptible to the infestation and provide a predictive distribution spread for potential and known Sahara mustard weed habitants. This chapter identifies the client and its problem. It discusses the proposed solution to address the problem.



**Figure 1-1: Map of Joshua Tree National Park, California**

## 1.1 Client

The client is JOTR National Park Service (NPS). The point of contact for this project is JOTR's Sean Murphy, a GIS specialist, NPS Ranger, and member of JOTR's Natural Resource Management team. JOTR comprises 794,000 acres with 585,000 of those acres designated as wilderness. Within JOTR boundaries are two deserts, two large ecosystems whose characteristics are determined primarily by elevation. Below 3000 feet, the Colorado Desert covers the eastern part of the park where the natural gardens of ocotillo and cholla cactus reside (see Figure 1-2). The western half, at a higher elevation, is the Mojave Desert with slightly more moisture and slightly cooler temperatures, and a special habitat of the Joshua tree (NPS U.S. Department of the Interior, 2012).

JOTR's mission is to manage, conserve, and protect the national park's natural habitat, archeological sites, historic structures, museums, and trails. The client helped define the problem and provided all data.



**Figure 1-2: Cholla Cactus Garden (NPS U.S. Department of the Interior, 2012)**

## 1.2 Problem Statement

JOTR needed to identify high risk locations of Sahara mustard weed infestation and to analyze potential weed distribution. With this knowledge, JOTR could then develop an effective control strategy plan for treatment prioritization and the efficient use of JOTR resources. Once these habitat locations were confirmed, JOTR needed to analyze the adjacent areas to determine the best approach where the native plants nearby would not be negatively impacted by the selected control method. After analyses, the JOTR staff

needed a map to help with orientation when the control treatment phase began. A GIS was the best application to help in the identification of current and future locations of the Sahara mustard weed habitat and map production. Benefits from the development of such a GIS for the JOTR staff were to be a better understanding of the weed's habitat, more direct management, and more efficient utilization of resources in the eradication of this plant.

### **1.3 Proposed Solution**

The proposed solution was the development of a GIS using ArcGIS. The GIS includes: a JOTR geodatabase for land assessment analysis, and a Predictive Species model. The geodatabase would store essential land data layers for analysis and aid in the development of actionable information.

#### **1.3.1 Goals and Objectives**

The goal of this project was to develop a GIS to support JOTR eradication's efforts against the Sahara mustard weed in JOTR. There were two objectives. The first was the identification of Sahara mustard weed habitat locations, the potential and spread distribution. The second objective was the development of a geodatabase for analysis and to store data.

#### **1.3.2 Scope**

The scope of this project was limited to the identification of the Sahara mustard weed in JOTR using ArcGIS as the developing environment and geodatabase management. The Sahara mustard weed is not the only invasive plants species in JOTR, but since it has high impact risk level, it was designated as the target species that would be analyzed using the GIS for the JOTR staff. It was hoped that successful implementation of this GIS could lead to addition of other invasive species models in the future. Project deliverables were a geodatabase, user document and Python scripts.

#### **1.3.3 Methods**

The first step of this project was to review and format the large amount of data provided by the client. All soil data come from the United States Department of Agriculture's (USDA) Natural Resources Conservation Service (NRCS). The complete Soil Survey Geographic (SSURGO) database for JOTR was downloaded from NRCS's Soil Data Mart website. Careful reviews of the SSURGO metadata and data model diagrams were required in order to create the database table in ArcGIS, as the database had over 60 tables. The project's database model was developed with contribution from the framework of ESRI's Biodiversity data model. The primary key entities were:

- Biodiversity unit – Invasive Species Type – Sahara mustard weed
- Observation – Recording of NEW weed sightings
- Distribution – weed spread: Known, Predicted, Potential, Historical
- Site – Geo-coordinates of KNOWN weed habitat locations
- Area – Area sections in JOTR

- Soil Type – JOTR soil type
- Trails Roads – in JOTR

Finally, guidance on the GIS data layer requirements for invasive weed modeling was researched from USDA's Forest Service Remote Sensing Application Center (USDA Forest Service Remote Sensing Applications Center, 2011).

## **1.4 Audience**

The target audience for this project report is the GIS faculty at the University of Redlands, the JOTR GIS professional staff and scientists, and the National Geospatial-Intelligence Agency (NGA).

## **1.5 Overview of the Rest of This Report**

The following chapters discuss the rest of this project. Chapter Two covers the research, background, and literature reviews done for the project. Chapter Three discusses the project's system design incorporating the client's strategic goals and requirements. Database design is explained in Chapter Four, describing the logical model in relationship diagrams. Project implementation is described in Chapter Five. Chapter Six and Seven discuss the GIS tool results, analysis, conclusions, and future work.

## **Chapter 2 – Background and Literature Review**

The goal of this project was to help JOTR find the Sahara mustard weed. The Sahara mustard is the number one invasive weed in the Southwest. This chapter covers the Sahara mustard's fact sheet and the literature review of the challenges of modeling species distribution when there is no concrete understanding of this invasive complex ecology. Today's computing strength provides accessibility to additional powerful multivariate statistical tools in GIS. These tools aided the development of variations or combinations of predictive distribution models that can analyze such factors as species management, conservation biology, biogeography, and climate change (Guisan & Zimmerman, N.E., 2000).

### **2.1 Invasive Sahara Mustard Weed**

Sahara mustard is native to North Africa, the Middle East, Pakistan, and the Mediterranean region of southern Europe (Minnich & Sanders, 2000). The first Sahara mustard report in California was in 1927, most likely brought over with the date palms imported into Coachella, CA from the Middle East (Halvorson & Guertin, 2003). By the 1970s, the weed had spread throughout the lowland deserts of northern Baja California, southeast California, and southwestern Arizona where much of the land consists of sandy soils – much like the Middle East.

Sahara mustard germinates, grows, and reproduces before local annual species, therefore claiming resources before competition occurs (Halvorson & Guertin, 2003). It flowers from late winter to early spring. This invasive weed is drought tolerant, fire adaptable, can self-pollinate, and a single plant can produce 750-900 seeds. In a hot, dry environment, the weed can break off from the soil base and roll as a tumble weed across the desert distributing seeds. Rain facilitates the production of a sticky gel covering the seed pod, permitting long distance dispersal of seeds by adhesion to animals, people, and even autos (Minnich & Sanders, 2000). There is no concrete information on the length of seed lifespan, although Minnich and Sanders (2000) speculate it may be years. It has been observed that temperatures in the low 20s successfully kill off many seeds (USGS Western Ecological Research Center, 2009). Sahara mustard is highly vulnerable to soil salinity and a poor competitor when native or other exotic plant densities are high.

The Sahara mustard is commonly found in open sandy places, sandy-gravelly washes, low dunes, interdune troughs, sandy alkaline flats, rocky slopes, and disturbed areas such as roadsides and open fields (Devender, 1997). In a 2006 study of spatial models for potential Sahara mustard distribution, Erick Sanchez-Flores (2007), found that the human-related factors strongly influenced the distribution of invasive species. Human-related variables include property type, trails, backcountry roads, highways, ranches, railroad, tourist interests, and small human settlements. Road networks were demonstrated to be the strongest predictor of presence. The study confirmed the hypothesis that more dynamic landscapes are prone to invasion by Sahara mustard.

The Sahara mustard has been observed to be the number one invasive weed with the impact to cause the most native ecological damage to the JOTR and Southwest desert regions. Desert ecosystems live on the extremes of viability – high spatial and temporal

variability of water – and are sensitive to disturbance that interconnected and complex ecological relationships are directly impacted (Sanchez-Flores, 2007). Once established in an ecosystem, the Sahara mustard alters competitive interactions, microclimate, native species' productivity, and the ecosystem fire regime. The weed's increasing plant cover smothers native plants and disrupts native animal habitats. It increases fuel loads in the desert, elevating the risk for fire, and threatens native species in areas where the species are not adapted to fire (Williams & Baruch, 2000).

The southwest states of Arizona, California, Nevada, and New Mexico have government councils, national parks, universities, and museums where the Sahara mustard is featured prominently on their invasive weed lists. In the California Exotic Pest Plant Council 1997 Symposium, six invasive plants were featured in the “Rogues Gallery of Exotics”, and the one lone invasive weed selected was Sahara mustard (Devender, 1997). The Anza-Borrego Desert State Park, along with other desert organizations and concerned citizens, formed a Sahara Mustard Weed Eradication Task Force (Commerce, 2010) for public awareness and eradication efforts. The University of California in 2010 developed the Sahara Mustard Consortium, which supports the scientific research and implementation of removal procedures, including biological control, aimed at reducing Sahara mustard in southwestern North America (McDonald, Chris; University of California Cooperative Extension, 2010). In 2009, Chevron Energy Solutions was awarded the contract to build the first solar energy project using photovoltaic technology on public lands in San Bernardino County, CA (U.S. Department of the Interior, 2009). The Bureau of Land Management (BLM) required that Chevron Energy Solutions develop and monitor a weed control plan during the entire construction phase of the project. BLM's main concern, documented in the control plan, was the Sahara mustard weed (Chevron Energy Solutions, 2010).

But after 40 years, with all the available resources, cooperation and knowledge on invasive plant species, ecologists have yet to identify the natural rules that govern Sahara mustard infestation (Sanchez-Flores, 2007). This makes it difficult to model the complex interrelations between the mustard weed and the ecosystem, and its distribution potential. This created an information gap on effective control methods and management strategies. The best eradication method to date is to hand pull the Sahara mustard with the roots before the seed dispersal phase. JOTR and the Morongo Basin Conservation Association (MBCA) annually organize “Hold the Mustard” weed pulling events in March to battle the spread.

The Sahara mustard displaces native plants and increases the potential for fire. The best control method is labor intensive and challenging because disturbing the soil causes dispersed seeds to spread. In the following section, developing a model for potential habitat distribution and the use of remotely sensed datasets are researched in the hunt for the Sahara mustard.

## **2.2 Predictive Habitat Distribution Models**

Ecological models generally have three properties: generality, reality, and precision (Guisan & Zimmerman, N.E., 2000). A model can have two but not all three properties. The mechanistic model is the most widely applied model for habitat suitability modeling. It is based on cause-effect relationships and uses an ordinary multiple regression, Generalized Linear Model (GLM), to predict the invasion rate. The mechanistic model

has a variation, called niche modeling, based on the understanding of how a species' properties interact with its surroundings to influence its fitness (Kearney, 2006). This model maps the derived niche to a real habitat capturing a fitness factor to predict the species distribution. One of the niche models, Ecological Niche Factor Analysis (ENFA), uses only species presence data, whereas the GLM requires both species presence and absence data. A comparison done for both models by the University of Lausanne (Hirzel, Helfer, & Metral, 2001) on simulated species was based on three scenarios: spreading, at equilibrium, and over abundant species. The development of a simulated species was essential to examine the various distribution patterns of biologically identical species in a mutual environment. The distribution was very sensitive to the data quality. The assessment demonstrated that the ENFA model worked as well as the GLM model. The GLM worked slightly better depending on the quality of the data but not on quantity of data. One essential item these mechanistic models and their variations cannot provide is the pattern of the invasive spread.

The need for predicting rates and patterns of alien plant spread initialized the formalization of a spatially explicit, individual rather than community-based simulation model that places greater emphasis on the plant-environment interaction (S.I. Higgins & Cowling, 1996). This model is structured to include the explicit modeled ecological factors that are major determinants of a targeted species' spread. Some examples of ecological factors can be fire survival, seed dispersal ability, climate change, and maturity rate. The study compared the mechanistic model and spatial explicit individual-based simulation model, and found the cause-effect model failed to mimic ecological processes and interactions, thereby reducing the model's predictive ability. The spatial simulation model's predictions were more sensitive and the resulting spatial data can be used by a GIS for pattern analysis. The performance of the spatial simulation model was greatly influenced by knowledge of the ecological process and the spatial scale of the processes. A strong collection of empirical data is critical to the success of this model.

A mechanistic version of the environment envelope model, which is based on calculating a minimal species-specific envelope in a multi-dimensional environmental space, uses only the lower limits of the direct and resource gradients (Prentice, Cramer, Harrison, Monserud, & Soloman, 1992). This approach predicts the predominant plant functional types by competition rather than by intolerance or environmental constraints.

The challenge was to select a realistic distribution model without the comprehensive understanding of the Sahara mustard's biological traits. Of the research presented thus far, not one is on invasive plant species in a desert environment. In Sanchez-Flores's study (2007) of using different variables intervening in the prediction of Sahara mustard spread, he utilized the genetic algorithm rule-set production (GARP) approach. The GARP method is based on a set of conditional rules in the form of "if" statements. These statements describe the ecological niches of the target species. The main advantage of GARP is the capability to apply different types of rules at once to explain complex non-linear relationships between the species occurrence and predictive variables (2007). The rule types are:

- Atomic – uses a single precondition variable
- Envelope – defines the environmental limit range of a species
- Logit – output is converted into a probability based logistic regression



GARP is based on presence-only data. But when the model dataset is a limited size and the predictor variables use different measurement scales, GARP was then demonstrated to be the best method for predicting species distribution based on point occurrences (Peterson, 1999). GARP was deemed the best approach for this project's predictive model.

### **2.3 Remote Sense Datasets for Habitat Suitability Models**

Remotely sensed datasets are being produced faster with lower costs and readily becoming available to the consumer. The use of these datasets is growing by ecological modelers and biogeographers for measuring the biophysical properties of the ecosystem, detecting environmental change, validating the accuracy of data, and in particular, modeling habitat suitability (Kerr & Ostrovsky, 2003). An assessment by a team of researchers at University of California, Davis used advanced remote sensing techniques that demonstrated hyperspectral and LiDAR datasets are sufficient to generate high-resolution habitat predictions for an invasive species since these techniques capture subtle variations in land coverage (Andrew & Ustin, 2009). Hyperspectral signatures of some plant species can be used to clarify the ecology of the species in the target habitat. With the growing use of remotely sensed datasets, the discoveries of both the positive and negative aspects are becoming clear. Research conducted by University of Massachusetts's Department of Environment Conservation (Bradley, et al., 2012) found the high potential for bias in the habitat suitability modeling with the inclusion of remote sensing variables as predictors if there was not a clear understanding of the ecological relationships in the model. It was decided that remote sensing datasets should characterize potential habitat rather than actual species distribution.

### **2.4 Summary**

The invasive plant species can be a more effective agent of significant ecological change than global warming (Sanchez-Flores, 2007). With this in mind and the accessibility of increasing computing power, choices and combination of species distribution models will continuously evolve. The modeler has to ultimately rely on the knowledge at hand and the data available to help provide guidance to an appropriate initial set of suitable models. There are notable knowledge gaps in the Sahara mustard's behavior and complex interactions with the desert landscape, but a model based on available knowledge and data can help close the gap.

## **Chapter 3 – Systems Analysis and Design**

This chapter presents an overview of the methods used to develop the GIS design which ultimately supported the client requirements and their strategic purpose – protecting and preserving JOTR National Park. Understanding the information products and analysis the client needed drove the formulation of system requirements – both functional and non-functional – that were incorporated into the blueprint of the final system design. The GIS toolset is comprised of two core components: a geodatabase for land assessment analysis and a weed distribution model. A project plan was developed to help identify and rank project tasks of the system design into a feasible timeline.

### **3.1 Problem Statement**

The Sahara mustard weed is the number one invasive weed threat in the Southwest low desert regions. It can alter the desert ecosystem for many of the native species plants and animals. The JOTR management staff needed GIS techniques to help identify, forecast, and track this desert invasive habitat in and around the JOTR National Park.

### **3.2 Requirements Analysis**

#### **3.2.1 Data Requirements**

Identification of required datasets was vital in providing the relevant information used for analysis and management planning. JOTR staff had a large amount of JOTR land and vegetation data but they were not formatted to usable GIS layers needed to perform any analysis or modeling. Taking guidance on weed management methods from the USDA's RSAC (USDA Forest Service Remote Sensing Applications Center, 2011), the first task was to identify the invasive species specific habitat requirements. Limited is known about specifics on Sahara mustard. Recommended GIS data layers from the literature for invasive weeds analysis were:

- Vegetation cover
- Elevation
- Disturbance areas – parking lots, visitor area, campgrounds, etc.
- Slope
- Climate data
- Soil type
- Hydrology
- Roads
- Trails

The other critical dataset was the current and past known Sahara mustard habitat locations throughout JOTR National Park.

### 3.2.2 System Functionality Requirements

Functional requirements guided the architecture design for the system, dictating the results or outputs of the system. Non-functional requirements defined the quality and usability features of the system. Because this was an initial GIS system for the JOTR staff to help identify, monitor, and analyze an invasive plant species in their workplace environment – the primary non-functional requirements were the system had to be easy to understand, easy to use on a daily basis, and easy to be expanded. Not all JOTR personnel who support the eradication efforts are GIS experienced; nonetheless they are required to collect data and import them into the geodatabase, do analysis, and participate in planning the annual JOTR Hold the Mustard events. There were two primary information products to be generated - identify the land areas most vulnerable for potential infestation and realistic predictive spread from those locations once habitat locations were discovered. Table 3-1 lists functional and non-functional requirements.

**Table 3-1: Function and Non-Functional Requirements**

Type	Requirement	Description
F	Land Assessment Analysis	Geodatabase for a land assessment analysis with JOTR land types
F	Predictive Spread Model	GIS model for predictive weed spread with weed location points
F	Geodatabase	Repository to capture, query, edit, analyze and geo-process JOTR data and weed points
NF	System Interface	Intuitive User interface to system tools
NF	System Ease of Use	Suitable for use by non-GIS personnel
NF	Documentation	Documents on GIS tools, metadata and geodatabase schema

Because a realistic Sahara mustard weed spread rate has not been validated, the user needs to be able to input different spread rates into the Predictive Spread Model. One of the system’s functions is to take in these different rates and perform the calculations to allow the user to assess the various spread results. Another requirement of system functionality was calculating the distance cost path layer used in predictive spread with user specified layers. These feature layers’ relative distance to mustard weed points were captured and used in developing the distance cost path. This functionality allows the user to develop distance cost paths based on the understanding of factors and to produce different spread rates. Overall the GIS functional requirements directed that the system be capable of taking in a different number of parameters for the tool to be flexible and expandable in predicting spread.

### 3.2.3 Non-functional Requirements

An evaluation of the current technology the client had in the workspace was conducted to assess technology requirements for the GIS project. The client had recently updated the hardware and software working environment, making it an ideal stage to host a new GIS. Additions to the JOTR management office were:

- One new GIS server: MS SQL Server 2012 R2
- Two new workstations: 32-Bit, Windows 7, two quad core processors @ 2.67 GHz, 4GB RAM
- ArcGIS Server Enterprise software
- ArcGIS version 10.1 Desktop Advanced software

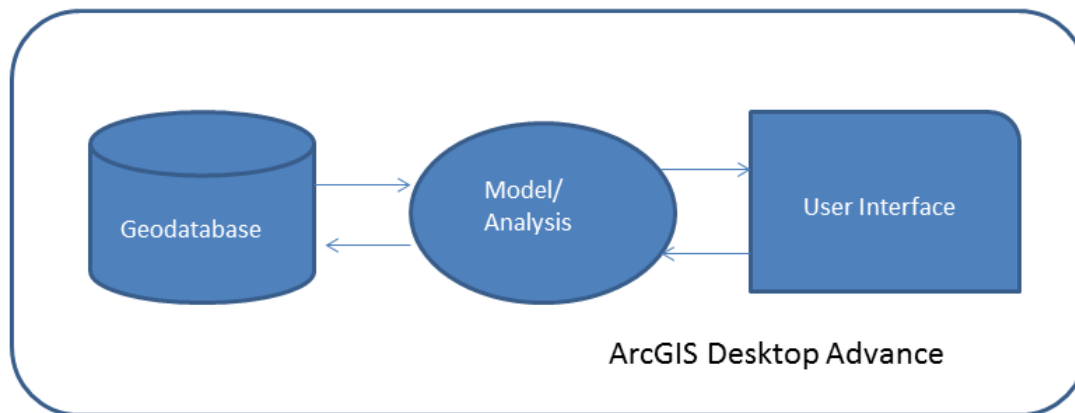
A technology review with the client provided additional information on equipment considerations in regard to:

- System Interface Requirements: This GIS was to be primarily a stand-alone system, no external link to data was required for processing. Imagery will be stored on an external hard drive. Generated informational products will be emailed or handed to consumers; no web hosting was required for this system.
- Communication Requirements: This GIS was to be accessed by two JOTR staff for park management decisions. Existing communication and network infrastructure met the requirements of this project.
- Hardware and Software Requirements: Needed for the development of this project's GIS, functional/non-functional requirements and information products were:
  - ArcMap version 10.1 for Desktop with Spatial Analyst and Spatial Statistics extensions
  - PythonWin: opensource scripting language
  - Microsoft Office: Word, Excel
  - No new hardware was required

After the evaluation, it was determined the client's existing technology infrastructure required no new additions or updates for this project.

### 3.3 System Design

After the assessment of the client and technology requirements, a system design was developed. Two JOTR personnel will regularly be accessing the toolset in support of their analysis. All data and processes are stored on the desktop system and web services are not required. An ArcGIS server was available, offering web based GIS capability, but the developed GIS toolset and the geodatabase were standalone entities, as the client requested. The heart of the system (see Figure 3-1) is the geodatabase, with all spatial data transformed into usable GIS layers that represent the desert landscape of JOTR. The tools are Pythonwin scripts, embedded into the ArcGIS in the form of a toolbox.



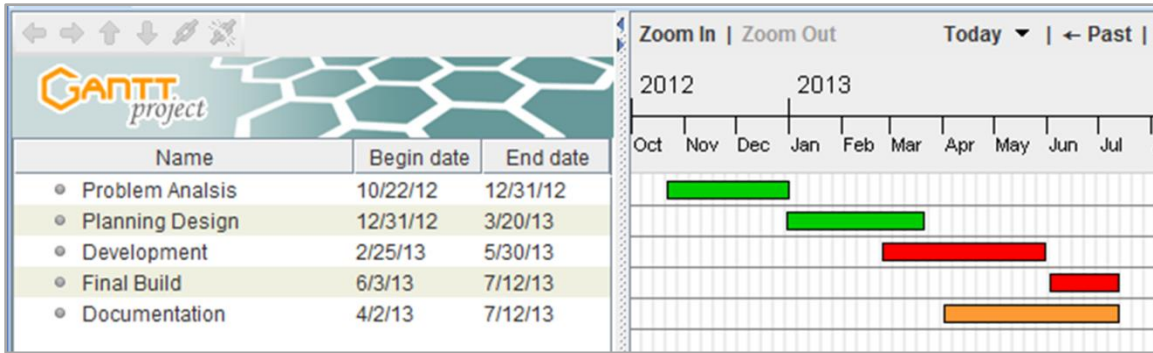
**Figure 3-1: System Design**

### 3.4 Project Plan

The initial project's phases were straightforward:

- **First Phase – Problem Analysis:** Gather and evaluate customer requirements, identifying the essential information and analysis products needed by client. Next analyze the current work flow process, to gain insight on how the project can be integrated into the current work cycle.
- **Second Phase – Planning design:** Determine the GIS data layers required and Sahara mustard biological traits to aid in environment suitability. Apply this knowledge to the design of the database. Evaluate researched predictive species distribution models to locate best fit model for Sahara mustard. Ultimately solicit guidance or input from the exotic species botanist and desert biologist on staff on these matters.
- **Third Phase – Construction of system with rapid prototyping:** With the identification of the prerequisite inputs for the desired outputs, build scripts to generate beta information products. Forward these products and scripts for evaluation by the customers on usability and performance.
- **Subsection of Third Phase: Customer Feed and Rework –** Initial prototypes of each functions will be sent to customer for feedback. Only valid requests will be applied.
- **Fourth Phase: Final Build and Testing**
- **Fifth Phase: MIP and User Documentation**

The time line estimated for the third phases with a complete working prototype was ten weeks. In reality, it took an additional four weeks. The Gantt chart (see Figure 3-2) shows this impacted phases starting and ending later in the project timeline.



**Figure 3-2: Project Gantt Chart**

The main factor that impacted the beginning three phases of the project plan was the project team not having the expertise to help develop an accurate invasive species data model and design a practical system in the time line proposed. Taking additional time for more research, the project team worked to understand the target invasive species and the environmental factors it favors. Analysis methods used by other land managers were researched that were related to invasive plant species control management. What was surprising during this time was the discovery of the knowledge gap on the Sahara mustard. Valid correlations have not been confirmed between any native species and this weed. There were no species spread models discovered for the Sahara mustard or any other desert plant invasive species. In regard to the system design, with so many unknown factors, different combinations of input parameters by the user were necessary to implement. The only site that offered any guidance on tracking weed invasive species was the USDA Forest Service; Salt Lake City, Utah office (USDA Forest Service Remote Sensing Applications Center, 2011). The site provided the framework and guidelines for land managers initiating weed management programs. The final model was developed based on the knowledge acquired from research and common knowledge of the project team.

### 3.5 Summary

The resulting down-to-the-basics developed system offers value and can be effective. The GIS system provided all the foundation of JOTR GIS layers for analyzing known and newly discovered Sahara mustard locations. The requirement analysis methods helped define requirements. The project timeline had to be altered when the project team needed additional time to gain invasive weed knowledge and customer requirements and desirable system outputs were not clear.

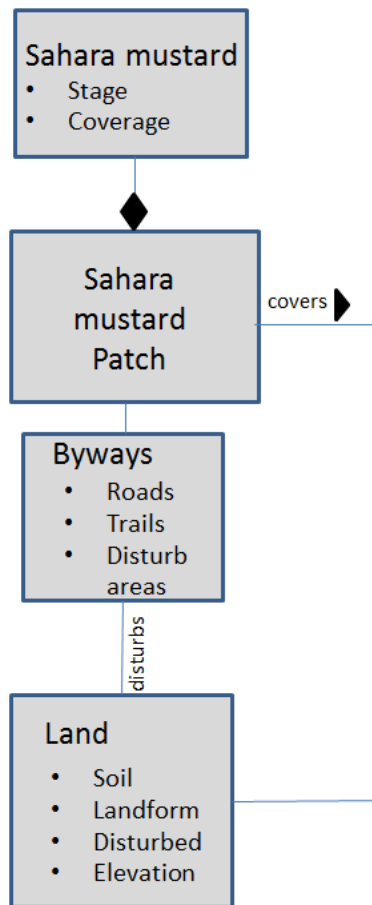


## Chapter 4 – Database Design

This chapter covers the design decisions of the database that geographically modeled the JOTR's environment and the Sahara mustard weed. Section 4.1 presents the conceptual model of the Sahara mustard weed attributes and relationships that influence its distribution. Section 4.2 describes the feature classes and tables that represent the necessary GIS data layers. The last three sections discuss the data sources, additional data collection and the data scrubbing required.

### 4.1 Conceptual Model

Human activity is the primary cause for accelerating the spread of the Sahara mustard in JOTR. The road network and trails are the byways of human activity. The factor that allows the Sahara mustard to exist is JOTR's suitable desert land. The biological factors, plant life phase, and land coverage size, of the weed patch observation add to the magnitude of its spread. These relationships are shown in the Sahara Mustard conceptual model presented in Figure 4-1.



**Figure 4-1: Conceptual Model of the Sahara Mustard Weed**



The figure shows that a single Sahara mustard plant may be part of a mustard weed patch and moves on byways to spread. Biological factors specify the quantity of individual mustard weed seeds that can potentially spread. Byways (road network, disturbed land) provide access to different areas of JOTR. Areas that have land factors that meet the suitability needs of the mustard weed are invaded. Land factors are soil types, elevation, landforms, and naturally disturbed areas (washes, alluvial fans, river bottoms). This conceptual model aided the next step of system development: the design of the logical data model and the GIS feature layers.

## 4.2 Logical Data Model

Presented here is the Sahara mustard logical data model: an abstract data structure based on the conceptual data model that identified important attributes and their relationships.

### 4.2.1 Sahara Mustard Weed Feature Class

This section describes the final developed main entity feature class Sahara Mustard and the major GIS land feature layers required for GIS analysis and weed spread modeling. Each entry in this feature class represents one observation of a mustard weed site. The Sahara mustard attribute table (see Table 4-1) exhibits attributes guided by the conceptual Sahara mustard model and with insight gained by the review of past JOTR mustard weed datasets. The two biological attributes were Stage and Coverage.

**Table 4-1: Sahara Mustard Feature Class Attributes**

Sahara Mustard Feature Class	
Date	Full date
Coverage	Low, Medium, High
Stage	Young, Mature, Dead
Collector	Name of data collector
Area	JOTR section name
Action	Inventory, Monitor, Treatment
Treatment	Manual, Chemical, Manual/Chemical
Comments	Any additional insight by collector
Year	Derive from date field

Stage indicates the stage of the plant’s life cycle: the rosette (young) stage; the mature stage where seedling generation is at its highest; or the dead stage. At the dead stage, the mustard weed had completed its life work. The seed pods have ripened and the seeds dispersed. The most likely scenario is that when this plant dies and dries up, the bush stem breaks off its base and tumbles across the desert dispersing seeds.

Coverage is the data detail on the size of the weed population extent at the time of observation (low, medium, high). This attribute field can be used to quickly assess where in JOTR are the largest weed infestations.

These two attributes, Stage and Coverage, provide valuable input in ranking and priority of which mustard weed sites need control treatment first based on threat impact.

Attribute fields Collector and Comments contain the collector's name and any insight by the observer at the time of the data collection. The field Area provides a more relative understanding of where in JOTR the weed site is located, for example, "Pinto Basin, mile number 11".

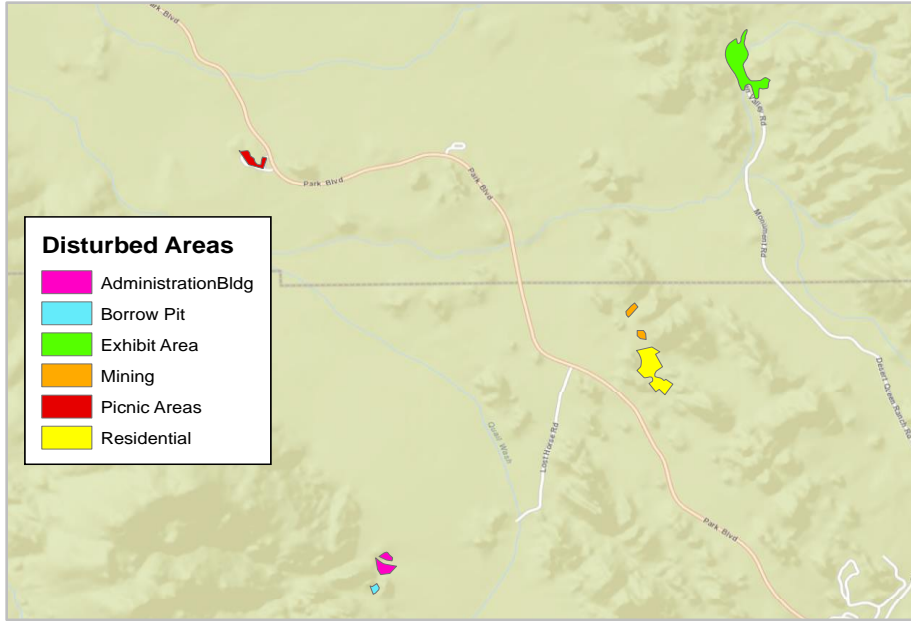
When field work is conducted on the Sahara mustard weed site the intent can be for three different purposes: inventory, monitoring, and treatment. These three action types will be the values of the attribute field Action. Inventory is the action of capturing new mustard weed locations. Monitoring is the assessment of prior observations. Treatment is the action of applying control treatment. A Treatment attribute field was added to collect the data on control method used on the mustard site. Chemical treatment is an option at JOTR but not widely used due to the potential negative impact on the native vegetation. The preferred treatment method is hand pulling. More data are required on the assessment and tracking of mustard weed sites by the JOTR staff to develop an effective weed control plan. Together all these attribute fields enable knowledge growth and provide the dataset to perform predictive analysis, weed inventory, population life stages review, growth assessment, rank, and manage treatment priorities.

#### **4.2.2 JOTR Land Layers**

GIS land layers were selected to best represent the environment where the Sahara mustard weed thrives. These layers included Soil, Dunes/Sand Flats, Landforms, Elevation, Vegetation classification, Slope, and Aspect.

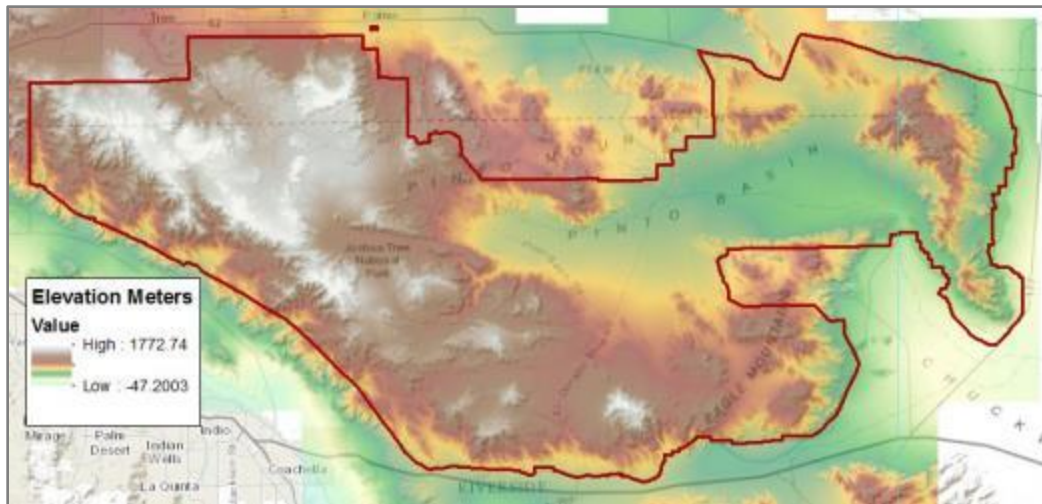
There were three soil feature classes representing the JOTR National Park land: Soil, Dunes/Sand Flats, and Landforms. Together these three feature classes offered a detailed perspective of the soil types in JOTR. The main feature class Soil described ten soil types. The Sahara mustard thrives in riparian habitats (river bottoms, arroyos, washes, etc.) (Devender, 1997). With this known fact, the second soil feature class Landform was added as a GIS land layer. It further described JOTR land by indicating the formation process type. This feature class contains 24 different feature types including: alluvial fans (leave deposits of gravel), fluvial terrace, erosional highland, and washes. The third soil feature layer was Dune/Sand Flats, which represented five soil types and was derived from a highly detailed vegetation feature class produced for JOTR by Aerial Information Systems (AIS), a geospatial firm based in Redlands, CA.

The Disturbed Land feature class was derived from the same JOTR vegetation class. Human activity creates disturbed soils in JOTR. This condition together with JOTR's desert climate is ideal for the mustard weed to prosper. These disturbed areas were created by activities such as the construction of buildings, road building, opening mines, and plowing land into fields. These areas can also be considered as a type of byway for the mustard weed to move. Figure 4-2 shows some disturbed area types on JOTR's Park Blvd.



**Figure 4-2: Close-up View of Disturbed Areas near JOTR's Park Blvd.**

The last land feature layer, Elevation (illustrated in Figure 4-3), can help identify potential mustard weed sites and aid in predicting its spread.

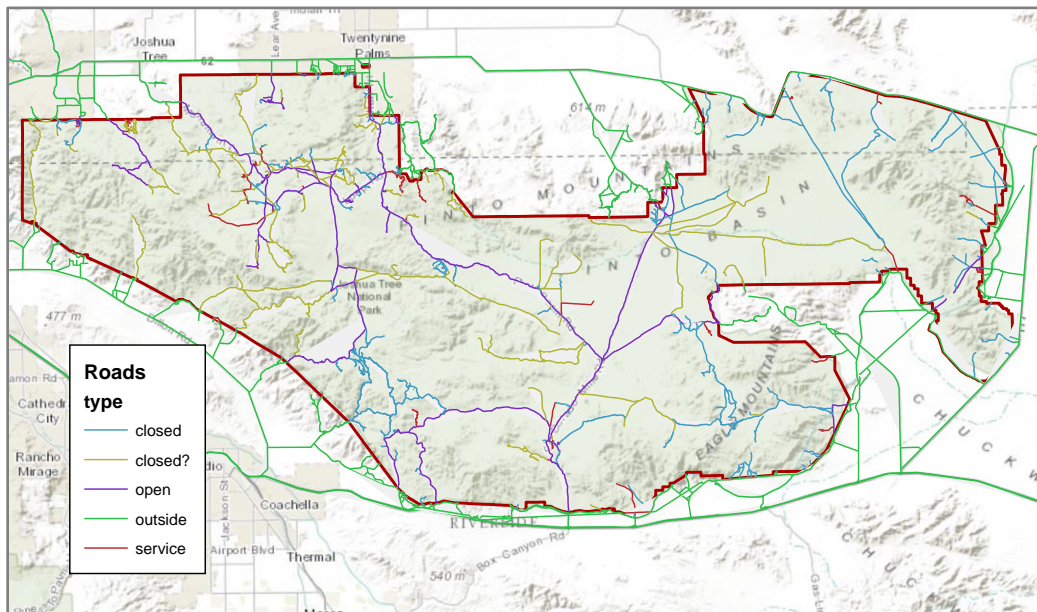


**Figure 4-3: Elevation Layer over JOTR**

### 4.2.3 Byway Layers

The Road feature layer shown in Figure 4-4 is considered the primary byway where the Sahara mustard weed is most commonly found and the main corridor for spread. The Road feature includes all the road types in JOTR: Open, Service, Outside, and Closed. The mustard weed usually sprouts on roads that are closed to the public; this may be a sign that the gravelly type soil on the roadside is still very favorable for weed growth.

The Roads layer is one of the main feature classes used in the predictive weed spread model.



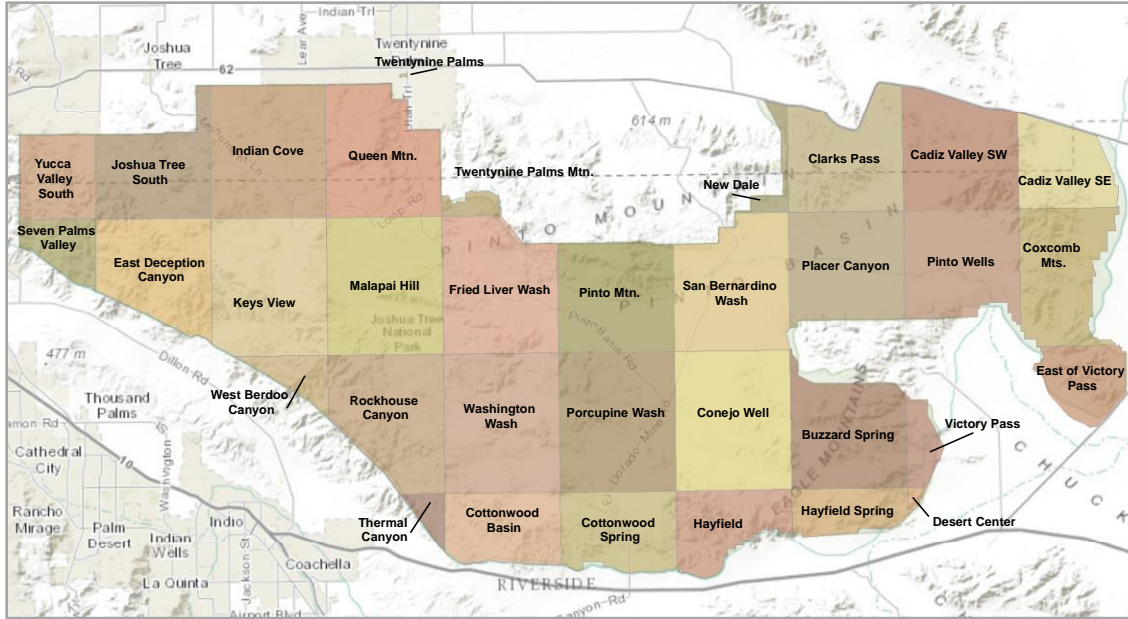
**Figure 4-4: JOTR Roads Layer**

#### 4.2.4 Other Useful Layers

These six feature layers: trails, vegetation classification, slope, aspect, SSURGO areas, and USGS section areas, were included in the database because they provide valuable data for analysis and in situational awareness. The JOTR Trail feature layer contains all hiking trails in JOTR. Hiking trails are the only way to access some of the national park's most beautiful and scenic areas. Mustard weed sites were not generally found on these trails but can still be a byway for weed spread. While there was not a confirmed correlation between certain JOTR native vegetation and the Sahara mustard weed, one study associates Sahara mustard weed patches with small leafed desert scrub and grassland in desert flat lowlands (Sanchez-Flores, 2007). The JOTR Vegetation classification feature class contains 67 different plant grouping and non-vegetation features. Determining the surrounding vegetation types near Sahara mustard observations may be useful in identifying potential JOTR areas susceptible to weed invasion.

The Aspect and Slope layers may offer insight when performing analysis over mustard weed locations. There were two layers illustrating JOTR park sections. One section layer was based on USGS topological 1:24,000 scale quadrangle tile names; the other was based on NCRS soil data survey (SSURGO) soil section names. The national park's total area is approximately 1,260 square miles. Associating a mustard weed site with a JOTR section provides orientation for the park staff. The JOTR staff uses the USGS naming conventions to refer to locations. Figure 4-5 shows the quadrangles used by JOTR.





**Figure 4-5: Overview of USGS Quadrangle Tile Names**

### 4.3 Data Source

The data used in this project came from five different sources. All past and current Sahara mustard weed points were collected solely by the JOTR staff. Examination of these datasets aided in the development of the logical data model. AIS produced the JOTR’s vegetation classification layer. The classification was initially done in the late 1990s while an accuracy assessment was performed in 2007 and 2008. After all updates and checks were completed, the vegetation feature class was finalized in 2010.

Together, Louisiana State University and U.S. Army Topographic Engineering Center developed the JOTR Landform feature class. Using a combination of space borne spectral scanners, air photo interpretation, and geological field techniques, they mapped the land forms, land composition, and earth materials of the Mojave Desert that stretched into California’s JOTR in 2000.

The USGS 1:24,000 scale topographic maps over JOTR were another data source used in this project. These topographic maps were the foundation for the digitization of all roads (open, closed, service) and trails in and around the park.

The final and fifth data source was the U.S. Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS) soil data. The JOTR soil data set was published in April 2012 and packaged as a Soil Survey Geographic (SSURGO) database. Soil data were represented in great detail and were developed by soil scientists as part of the National Cooperative Soil Survey.

Three of the five data sources were from either state or federal agencies which provided an extensive amount of metadata. The SSURGO’s database was comprised of six spatial components and over 60 tabular components. Documentation included two large relational diagrams covering all feature (spatial) layers and data tables. A separate SSURGO report detailed the entire structure for each data table and its attributes.

AIS developed the vegetation feature layers for JOTR and the metadata provided was minimal but sufficient. JOTR conducted a field accuracy assessment after receiving the GIS layers from AIS and validated their results. AIS’s website provides an overview of the vegetation classification project for JOTR.

#### 4.4 Data Collection Methods

All required data were provided by the client to implement this project. There was no additional data collection within this project.

#### 4.5 Data Scrubbing and Loading

The Sahara mustard weed data provided by the client was collected by different JOTR teams in the last few years. Closer examination of the data showed the main method of collection was using a GPS device, recording the location and number of weed observations (presence only) while driving down JOTR road ways. But two datasets, year 2008 and 2009, show the collection points were equally distanced from one another. The collection method here was presence/no presence data, where a set of locations were surveyed and recorded with either a weed presence or not. From these two datasets, locations with confirmed sightings will need to be extracted and imported into the appropriate dataset for analysis.

Most of the effort on data preparation went toward the SSURGO dataset which was in a relational format. The tabular data came in a comma-delimited text file, ready to be imported into the geodatabase and used for querying, analysis, and reporting. Each of the tables and spatial data had the same attribute field set as a primary key for indexing. This key allowed these related entities to be joined for access to all the record data. Unfortunately, most table names and their attributes were in encoded data or numeric values, making it difficult to use for identification and to be joined with related tables. SSURGO documentation was the essential road map. It provided information to identify and understand the meanings of file names, table names, and attribute names. Table 4-2 shows the metadata used to interpret the data files.

**Table 4-2: SSURGO Prefixes Identify the Referred Spatial Entity**

File Name Prefix	Spatial Entity
soilsa_a	soil survey area boundary polygon(s)
soilmu_a	map unit boundary polygons
solimu_l	line map units
soilmu_p	point map units
soilsf_l	line spot features
soilsf_p	point spot features
soilsf_t	spot feature description

By using the SSURGO data model diagram, the correct table Mapunit and spatial entity 'soilmu\_a' (MUPOLYGON), and their primary key 'mukey' were identified. These two tables were then joined and the soil data were extracted into a new table developed as the main Soils feature class. This feature class originally had three attributes: soil type, elevation data, and JOTR section name, grouped into a single attribute text field. String manipulations through the ArcMap Selection tool enabled the separation of each attribute into its own field. This formatting effort made the soils usable as a feature layer for analysis.

## **4.6 Summary**

Understanding how human activity, land factors, and biological traits of the Sahara mustard weed interacted in its spread were crucial in the design of the data model, geodatabase, and feature classes. The logical data design incorporated this understanding, setting a data collection standard for future mustard weed observations. With the Sahara Mustard weed feature class finalized, the environmental and human factor feature classes this invasive plant needed to exist and spread in JOTR were derived and imported into a geodatabase. Finally, an overview of data sources provided the validity of the GIS feature layers JOTR users can be confident in their analysis with the geodatabase and feature layers designs established.

## Chapter 5 – Implementation

After the planning and design phase, the two key components of this GIS were ready for construction. The components were a geodatabase for data storage and land assessment analysis, and a predictive spread model. This chapter describes how these elements were created and some of the challenges encountered during their development. ArcMap 10.1 was the main development software program used to construct and manage all three. The scripts that were developed for some feature layer processing were written in Python using the PythonWin developing environment. ArcMap models were created with ModelBuilder. All scripts and models are stored in a toolbox within the geodatabase for user access.

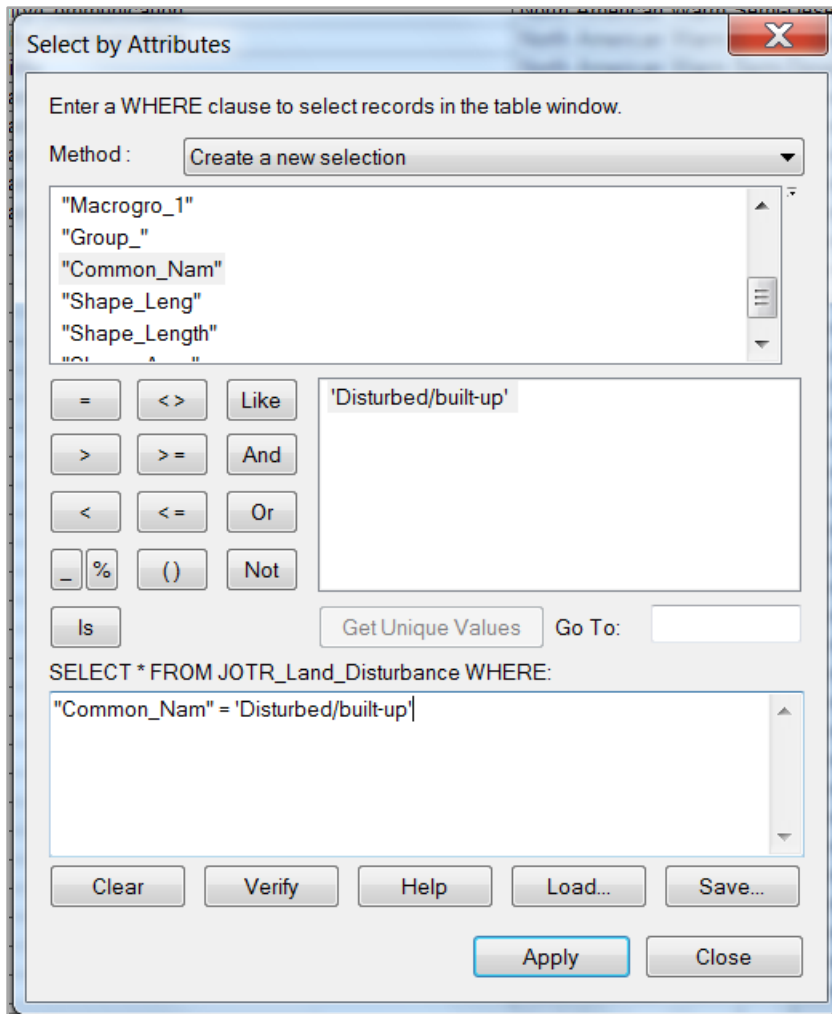
### 5.1 Geodatabase

Invasive species research provided the roadmap of information needed to track and assess the Sahara Mustard weed in JOTR. The mission after database design was to find high quality and reliable data and feature layers that could produce valid results from the planned GIS tools.

The JOTR client provided two gigabytes of geospatial data, including six geodatabases of data and an additional 80 feature layers outside the geodatabase. The geospatial data came from four different sources with most of the feature classes having metadata. A review and assessment of each geodatabase and the features layers were conducted to determine if they could or should be included in this project's geodatabase. A few feature layers that could have been used in this project were found without associated metadata to help interpret the information; in the end they were excluded.

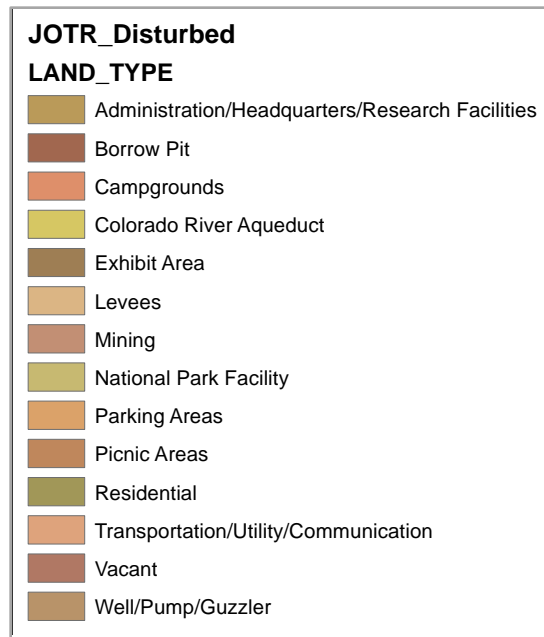
Three soil feature classes: Soils, Dunes/Sand Flats, and Landforms, were incorporated into the geodatabase; and they were collected from SSURGO, University of Louisiana/ U.S. Army Topographic Engineering Center, and Aerial Information Systems (AIS) respectively. All three datasets were very detailed. The Landform feature class had 24 feature types and the Soils feature class had eleven feature types. The JOTR vegetation classification feature class by AIS had 16,499 features categorized into 78 different feature types. It was from the Vegetation classification feature class that the Disturbed Areas and Dunes/Sand Flats feature layers were compiled. ArcMap's Selection tool (Figure 5-1) was used to select and extract each of these two feature types using an SQL statement.





**Figure 5-1: Selection Tool and SQL Used to Extract Disturbed Areas Features**

The land\_type attribute field in both resulted feature classes provides further classification information. For example, Disturbed Areas was classified into 14 subtypes as present in Figure 5-2.



**Figure 5-2: Disturbed Areas Types**

Three additional land feature layers, vegetation, slope, and aspect were provided in this project. There were no scientific facts on the correlations between these land factors and Sahara mustard spread and existence but these feature layers may provide valuable insight during analysis.

Also included in the GIS layers was the Digital Elevation Model (DEM) at five meter resolution. This layer was generated using Interferometric Synthetic Aperture Radar (IFSAR) and managed by National Oceanic and Atmospheric Administration (NOAA). The Roads and Trails feature layers were digitized by JOTR’s Sean Murphy based on USGS topographic maps. The two JOTR area section names feature layers, one based on SSURGO data and the other on USGS quadrangle tile names, were developed in this project. The JOTR MaskEnvelope polygon feature created for this project was used to define study area boundaries. The final feature class Brassica (from the Sahara mustard weed Latin name *Brassica Tournefortii*) was derived from the Exotics geodatabase provided by the client. This geodatabase includes all data of all invasive plants plaguing JTOR National Park. For this project, only the Sahara mustard weed (*Brassica tournefortii*) data were extracted and imported into the project database using the developed Sahara mustard logical data model. Data from different years had to be imported separately since each year’s data collection was formatted differently. Table 5-1 lists the compiled feature classes for this project.

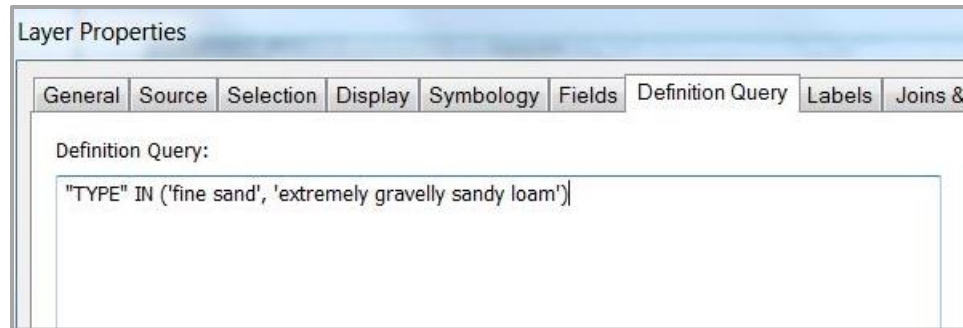
**Table 5-1: Feature Classes in the Project Database**

Finalized JOTR Feature Classes	
JOTR_Soils	JOTR_Area_SSURGO
JOTR_Landforms	JOTR_DEM_5m
JOTR_Dunes_SandFlats	JOTR_AllRoads
JOTR_boundary	JOTR_InRoads
JOTR_Area_QUAD	JOTR_Trails
JOTR_Disturbed_Area	JOTR_Veg_Class
JOTR_MaskEnvelope	Brassica (Sahara Mustard)
JOTR_Aspect	JOTR_slope

## 5.2 Land Assessment using the Geodatabase

There were many unknowns regarding the Sahara mustard complex inter-relationship with the JOTR desert environment so the capabilities of using different combinations of the feature layers were required for the land assessment analysis. Roads had been confirmed by experts as the human activity factor that is a good predictor for mustard weed invasion; but they were not used in the analysis for land suitability. Only the natural land factors that provide suitability for the Sahara mustard should be considered. The predictive spread model incorporated the human factors layers including roads.

With the soil feature classes having so many types, an SQL statement and an extraction script were developed to extract preferable soil types for Sahara mustard. The SQL statement was used for temporary data views using the Definition Query in a feature layer's property. This statement was defined under Definition Query from a feature layer's property. Figure 5-3 shows an example of such a SQL query statement. The feature layer Soils and the types of features to be extracted were fine sand and extremely gravelly sandy loam. TYPE is the attribute field. Query Builder can be used to build an SQL string with the correctly spelled values. The SQL statement follows a format: "FIELDNAME" IN ('value name1', 'value name2', 'value name3'). This query will filter out all records not included in the parentheses and ArcMap will only display the specified feature types in the parenthesis. This query will stay in the properties' definition query until deleted or changed by a user.



**Figure 5-3: Definition Query of the SQL Statement for display**

To create a new soil feature class for selected soil feature types, a Python script was developed called *SelectSoils* and imported into an ArcMap toolbox. A user will be able to select one or more soil subtypes to be extracted from the feature class using an interface that has a multivalued field. One of the challenges faced here during implementation testing, was an error in ArcMap's multivalued field input when it forwarded the soil type string parameters in an incomplete format, making some of the desired soil types missing from the resulted feature layer. To solve this issue, a Python function (see Appendix A) was developed to rebuild the entire incoming string parameter value list with correct SQL formatting. This function executes on the transmitted SQL string before it is used as an input parameter for creating the new feature layer. This same string function was embedded in the feature extraction script for the Dune & Sand Flats layer called *SelectDunes*.

With the extraction scripts and definition queries completed, the last layer option for this tool is reclassifying the elevation layer into different classes specified by a user. This can be done by using ArcMap's tool *Reclassify*. After all desired soil and elevations feature layers are created; the user can begin the analysis for land assessment. A use case is presented in Chapter 6.

### 5.3 Predictive Weed Spread Model

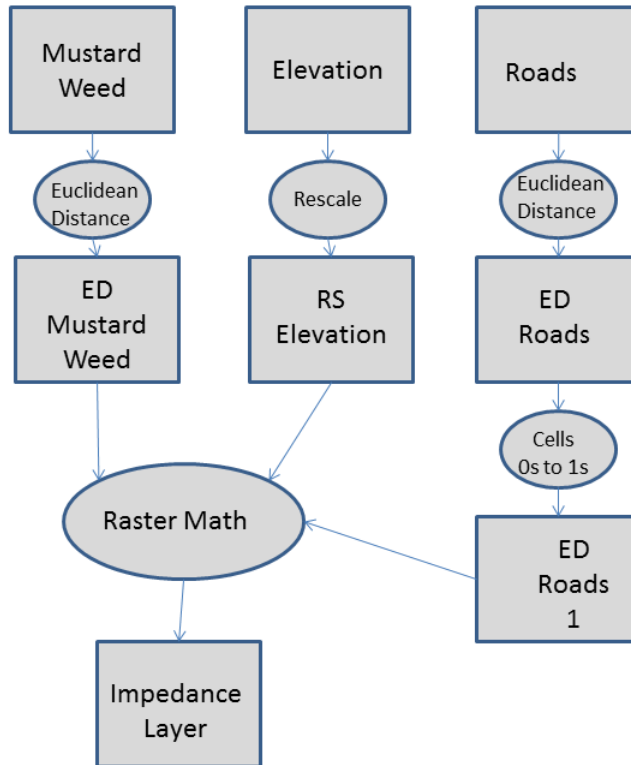
The Sahara mustard weed takes advantage of increasing human movement in JOTR, finding new opportunities to move into new desert landscape. Unfortunately there are so few known mustard weed factors to allow for reliable predictive spread. One acknowledged fact is that roads and trails help accelerate mustard weed spread (Sanchez-Flores, 2007). This may explain why over 95% of the Sahara mustard weed dataset provided by the client were located on the roadsides of JOTR. Another fact accepted by experts, the mustard weeds were commonly found at lower elevations (Devender, 1997). These specifics guided the methodology to develop a predictive mustard weed spread based on the weed observation proximity to road features with the factor of elevation.

The four data inputs required to implement this method were:

- Sahara mustard weed point observations
- Roads inside and outside of JOTR
- Elevation

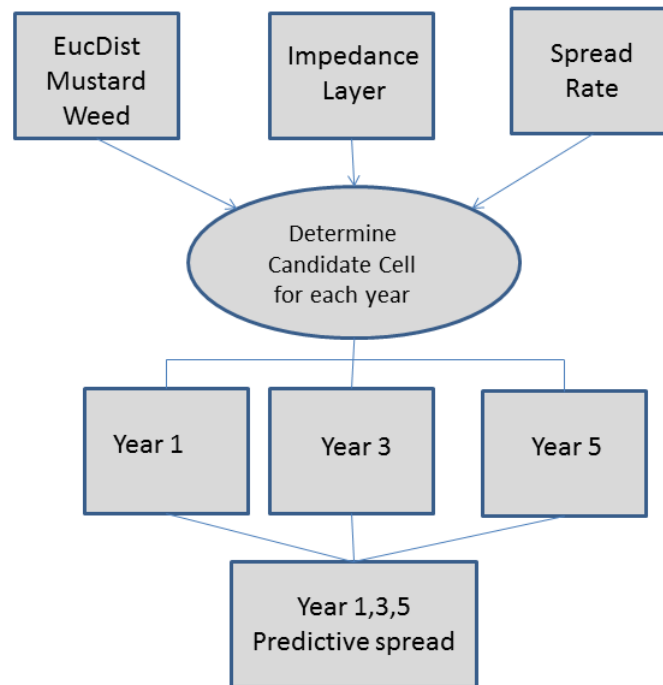
- Weed spread rate

The concept was to develop an impedance surface layer by using raster math algebra with these three features: mustard weed points, roads, and elevation. The impedance surface will help quantify how fast a weed spreads. Figure 5-4 shows the data flow for this phase of the implementation.



**Figure 5-4: Data Flow of First Phase of Predictive Spread Implementation**

Each layer's cell will have an impedance value, and cells with lower impedance will be identified as ideal routes for spreading. In the second part of the implementation, these impedance layers cells will be evaluated against an established threshold for a given year's spread rate. This analysis will limit the spread to cells with impedance smaller than the threshold. The final step in this method is to apply the spread rate to define the thresholds to predict weed spread for 1, 3, and 5 years. The data flow for this second phase of the implementation is shown in Figure 5-5.

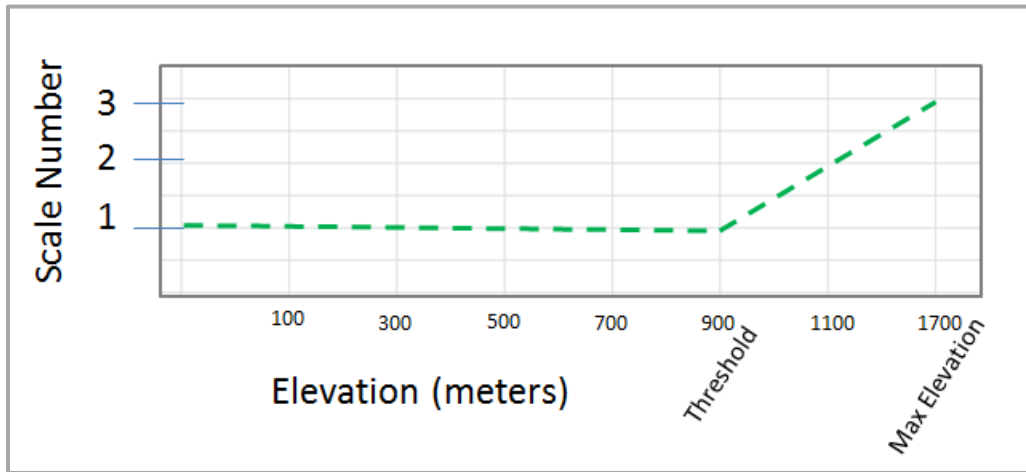


**Figure 5-5: Data Flow for Second Phase of the Predictive Spread Implementation**

The initial step of the implementation is setting the ArcMap’s environment to the cell size of 30 meters. The second step was to prepare the road and elevation layers for input. Since human factors are known to strongly influence the probability of Sahara mustard invasion, as an option the road layer can be merged with the trails layer and/or the disturbed land layers. A Python script was developed to provide this capability. Users can access this tool by selecting the SelectByways script in the toolbox. This tool also provides the additional option of creating a subset of the JOTR road types for input. The resulting feature layer will have all the specified human factors set by the user to contribute to the generation of the impedance surface layer.

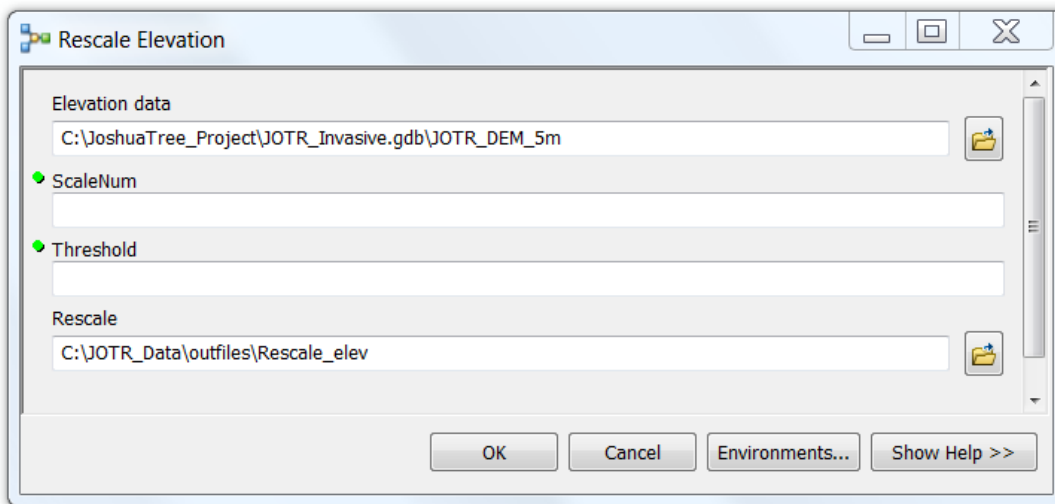
The input elevation layer is required to be rescaled by a user before using as an input file to the predictive spread model. The Sahara mustard weed prefers a certain elevation range to exist and spread, the maximum value of this range will be the threshold value. The difference between reclassifying versus rescaling elevation is the scale number applied to represent the elevation values. In reclassification, elevation values are sliced into intervals set by the user. All elevation values in an interval will be classified with the same value, a whole number. The first scale interval represents the most ideal elevation for the invasive plant species, the value will be 1. As the scale number rises in the reclassified elevation, the impedance factor rises for weed spreading. The impedance factor will have a constant value in the entire range of a specified scale. The results will be a noticeable jump in the impedance factor when going from one scale to the next. In using the rescale method, the first scale is also determined by the threshold value. Beginning from the minimum elevation value to the elevation threshold value, this scale

will be represented by the value 1, as like the reclassification method. But for the range beginning from the threshold elevation value to the maximum elevation value, the scale will begin from 1.0 and gradually rise to the max scale number set by the user (see Figure 5-6). The scale values will be in decimal numbers. The rescale elevation will generate an impedance factor layer whose values also will gradually rise with the elevation.



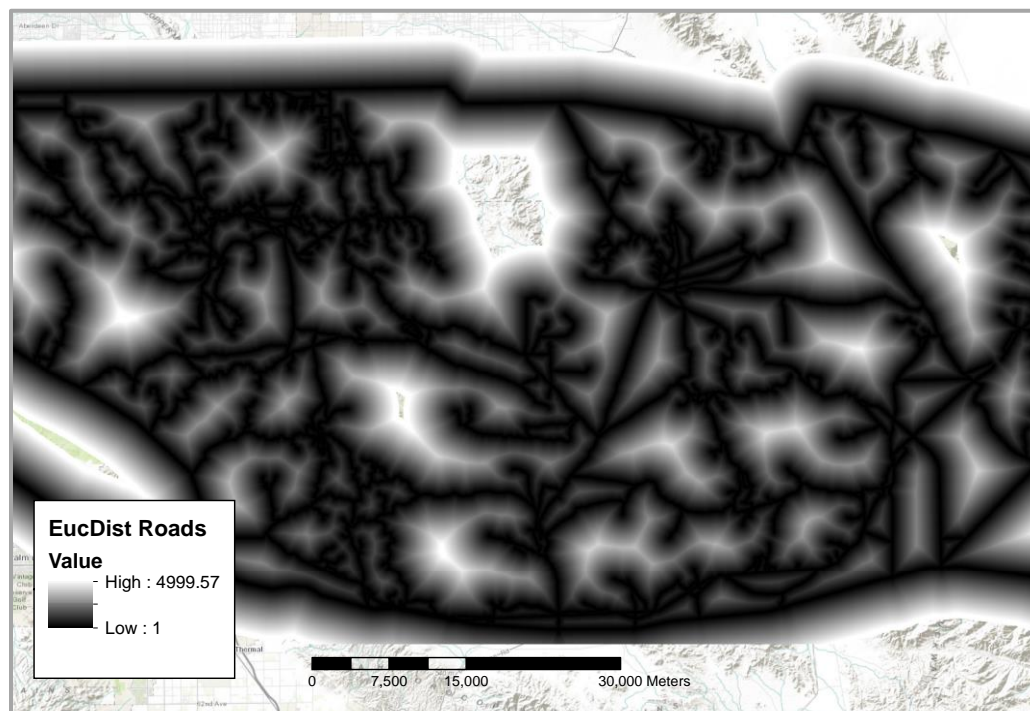
**Figure 5-6: Scale Values from Threshold to Maximum Scale Number**

The user can apply this method by using the developed ModelBuilder tool named Rescale Elevation (Figure 5-7). Inputs are the elevation file, the threshold value and the number of scales desired. Another input is the maximum elevation of the input file. ArcPy, a library that allows access to geoprocessing tools, was used to extract this internally, making this a standalone tool to work with any raster elevation data. Appendix B depicts the math equation and ModelBuilder data diagram.



**Figure 5-7: Rescale Elevation Tool Interface**

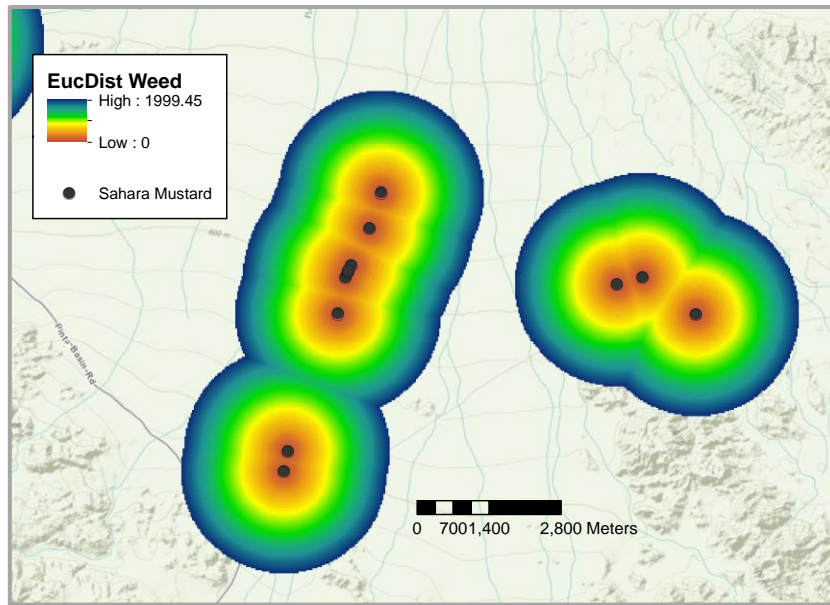
After the preparation of input feature files, the next step is to generate the impedance surface layer. Euclidean distances from the mustard weed points and road feature layers are calculated using ArcMap's Spatial Analyst Euclidean Distance tool. This tool calculates the shortest distance from a point to a feature such as a road. Euclidean distance results will be in raster format. The maximum Euclidean distance set for mustard weed points was 2000 meters and for roads 5000 meters. The resulting Euclidean distance layer for roads shown in Figure 5-8 presents roads in the color of black, and as the color of the roads fades to grey, the cells move outward while their values increase. Actual road locations in this Euclidean distance layer were represented by a 0 value, because this layer will be used as cost input, hence all values must be greater than 0; there are no zero costs in this method. To calculate a viable impedance surface layer, the next step in the implementation process was to recode the road cell values from 0 to 1. If zero cost values were desired to be used as cost inputs, these values would be updated to a small number, for example 0.01, for the multiplication process.



**Figure 5-8: Roads Euclidean Distance**

The Euclidean distance layer to Sahara mustard weeds is shown in Figure 5-9 with a color ramp to better perceive the distances.



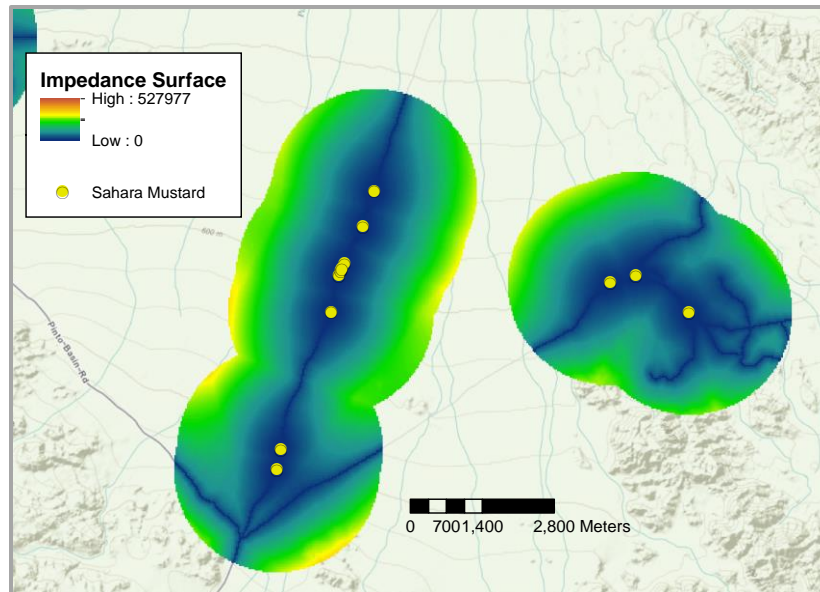


**Figure 5-9: Sahara Mustard Weed Euclidean Distances**

Map algebra was applied to these three layers to calculate the impedance surface. The Road Euclidean distance layer and re-scaled elevation layer were applied as weights to the Euclidean distance to weeds. To produce the resulting impedance layer with smaller values but with correct distance scale ratio, the square root of the Road Euclidean distances were used. The elevation re-scaled cell values were factored in by multiplication. The math algebra equation was:

$$\text{EuclideanDist\_weed} * \text{square root (EuclideanDist\_road)} * \text{Rescale elevation}$$

Figure 5-10 shows a close up view of a resulted impedance surface layer with weed points shown in yellow.



**Figure 5-10: Example of a Resulting Impedance Surface Layer**

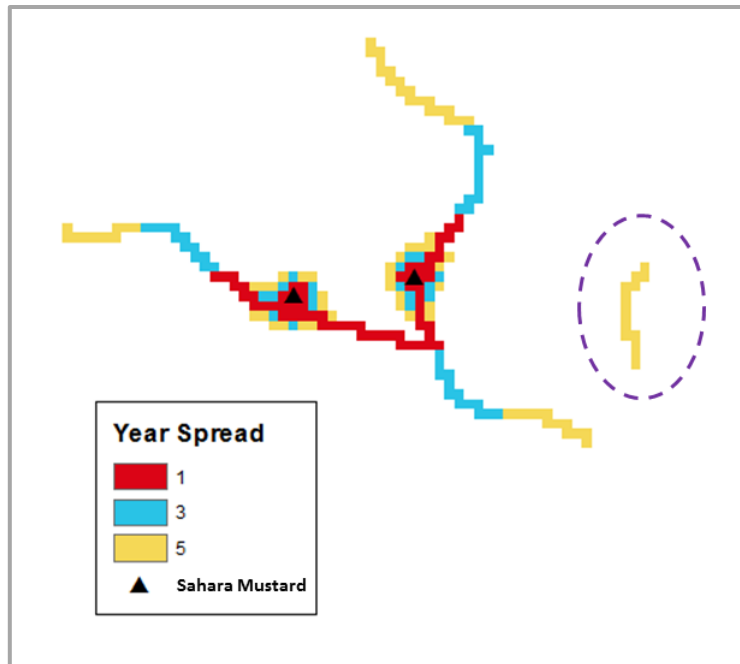
An ArcGIS model was developed to generate the impedance surface using the map algebra discussed earlier. The model is named PredSpread and listed in the toolbox.

The second part of the predictive spread model was developed in Python. The open-source add-on module of Numpy was used to process raster data. Numpy provides common mathematical and numerical routines in pre-compiled functions. The input for this implementation was:

- Impedance surface layer
- Sahara mustard weed Euclidean distance
- Yearly spread rate

The model predicts spreads for one year, three years, and five years. The yearly spread rate is multiplied by the number of years to establish the respective maximum spread distance. A threshold was used to identify the candidate cells for spread.

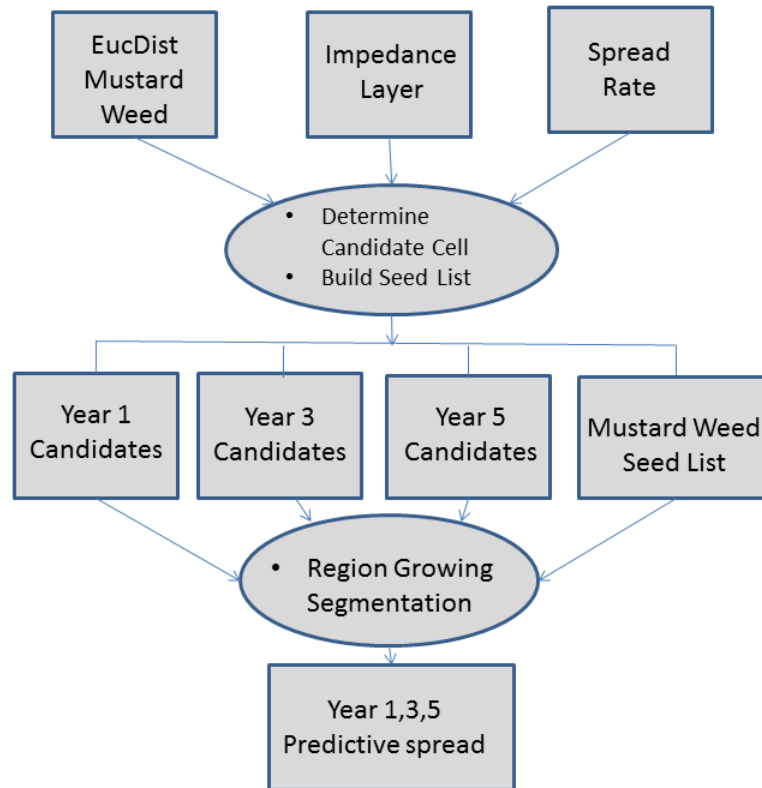
The process identifies cells in the impedance surface with values smaller than or equal to a threshold. Extracted cells are stored in Numpy arrays. Each year has an array to store the candidate spread cells. After all processing is completed; arrays will be output to a raster file. Examining the raster file showed some extracted cells did not originate from a mustard weed point as highlighted in the dashed ellipse in Figure 5-11. These were cells that met the impedance threshold criteria. Because we want to model the potential spread from a weed point source, the unconnected cells would have to be identified and removed. Only cells connected to a weed point will be kept as predicted spread.



**Figure 5-11: Unconnected Cells (i.e., dash lines) within Range of a Weed Point**

To solve this issue of candidate cells not connected to a mustard weed point, a technique based on the region growing image segmentation approach was used. Region growing uses seed points, which in this case were the Sahara mustard points, as the originating source to find connected neighbors with the same value. These connected cells were extracted as the valid weed spread locations. The region growing process was implemented using Python script (see Appendix C).

Each weed point is used as a seed in region growing process to identify valid spread locations. The extracted features by year were then processed out to the raster output file. The produced raster file will only have numeric values of 0, 1, 3, and 5, only. The value 0 will be for NoData cells and other cell values represent the weed spread locations for over one year, three years, and five years. Figure 5-12 shows the revised data flow for this second phase of the implementation process.



**Figure 5-12: Revised Data Flow for Second Phase of Predictive Spread**

## 5.4 Summary

This chapter described the implementation of the two main elements of this project: the geodatabase for data and land assessment analysis, and the predictive spread model. A region growing technique was used to solve an issue in the predictive spread model.

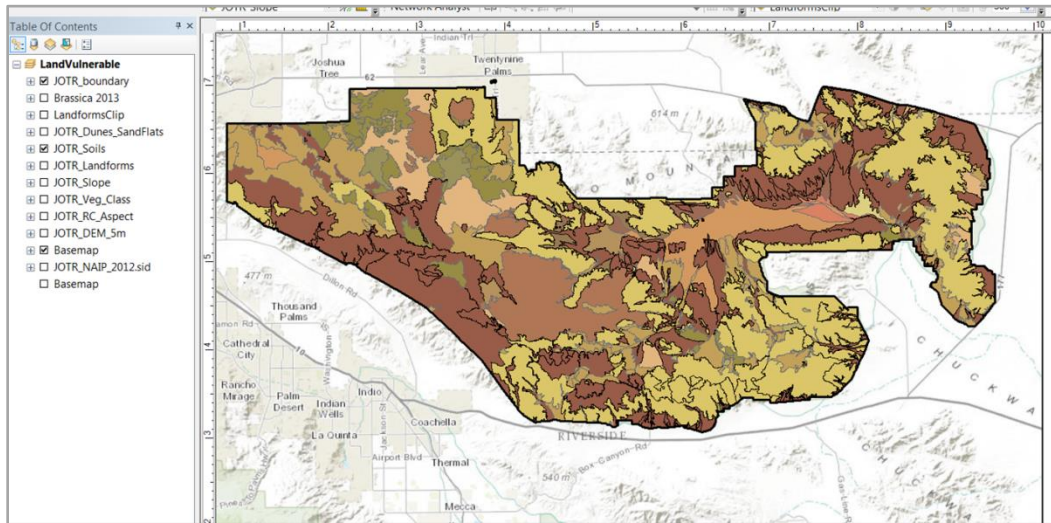


## Chapter 6 – Results and Analysis

The goal of these tools was to help the JOTR staff to identify and track the invasive Sahara mustard weed for eradication. A land assessment analysis was performed with a soil and elevation test case. Then, the outcomes of the predictive spread tool were examined after testing was performed with different spread rate parameters.

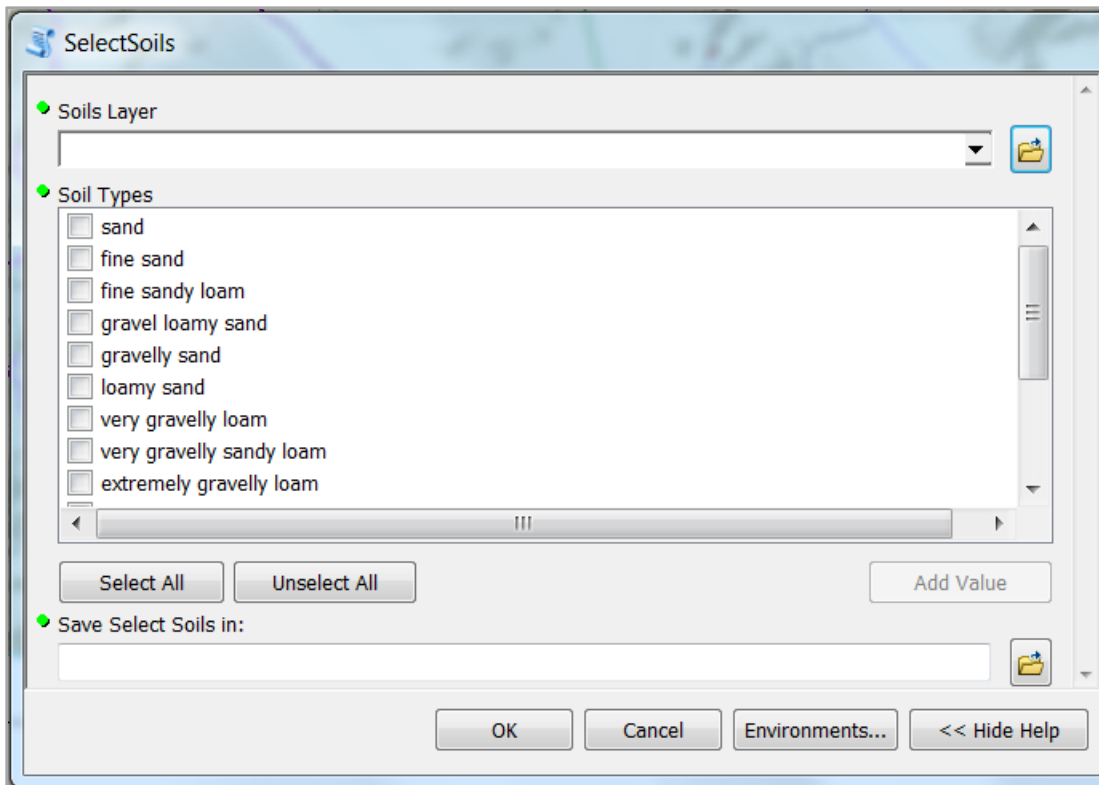
### 6.1 Sample Land Assessment Analysis

Users specify JOTR land features that best represent the specific habitat requirements of the Sahara mustard to be used to identify vulnerable areas to weed invasion. Figure 6.1 shows the interface to access the land feature layers for analysis. The two main land factors in this tool are Soil Types and Elevation. Elevation and sandy soils have been deemed by expert opinion as the key factors in identifying the weed’s habitat (Devender, 1997). The soil class Gravelly Sand was selected and extracted from the Soil and the Landform layers for analysis.



**Figure 6-1: Interface for Feature Layers for Land Assessment Analysis**

Figure 6-2 shows the user interface for the developed tool, SelectSoils, to extract desired soil subtypes from the Soils feature class. A user has the option to develop a SQL query instead of extracting feature layers by accessing the layer’s properties.

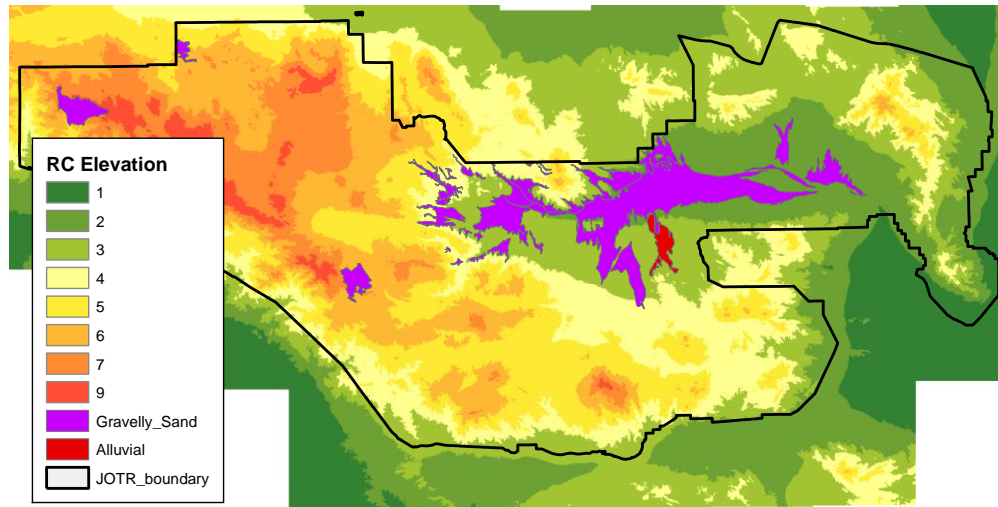


**Figure 6-2: User Interface of SelectSoils Tool**

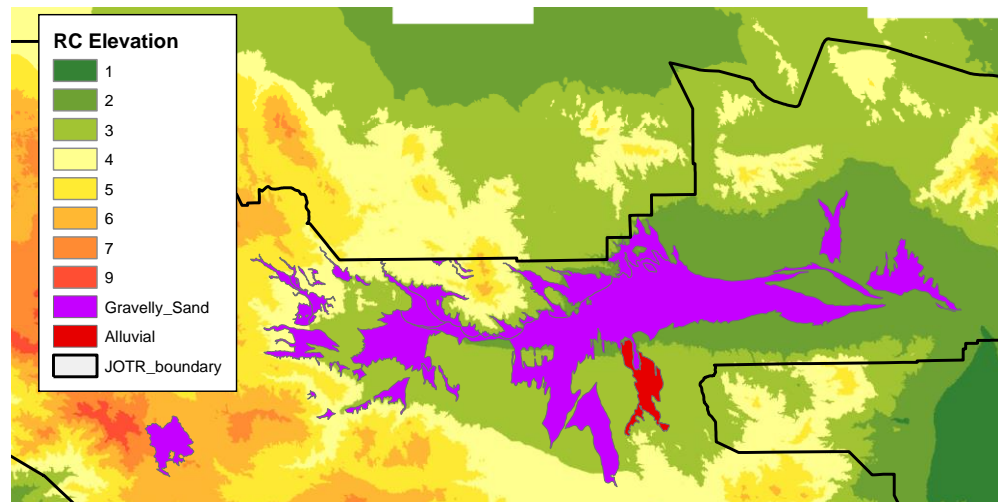
After desired soil types are extracted, the elevation layer is reclassified to represent the ranges of elevation by utilizing ArcMap's Reclassify tool. For this example, the elevation was reclassified into 8 classes:

- < 300 = 1
- 300 – 500 = 2
- 500 to 700 = 3
- 700 to 900 = 4
- 900 – 1100 = 5
- 1100 – 1300 = 6
- 1300 – 1500 = 7
- > 1500 = 8

Class 1, 2 and 3 represent the suitable elevation for the mustard weed and are green colored. Overlaying the extracted soil layers and the reclassified elevation layer shows the potential suitable areas for Sahara mustard weed as shown in Figure 6-3. Figure 6-4 shows a closer view of the potentially high risk areas for Sahara mustard invasion.



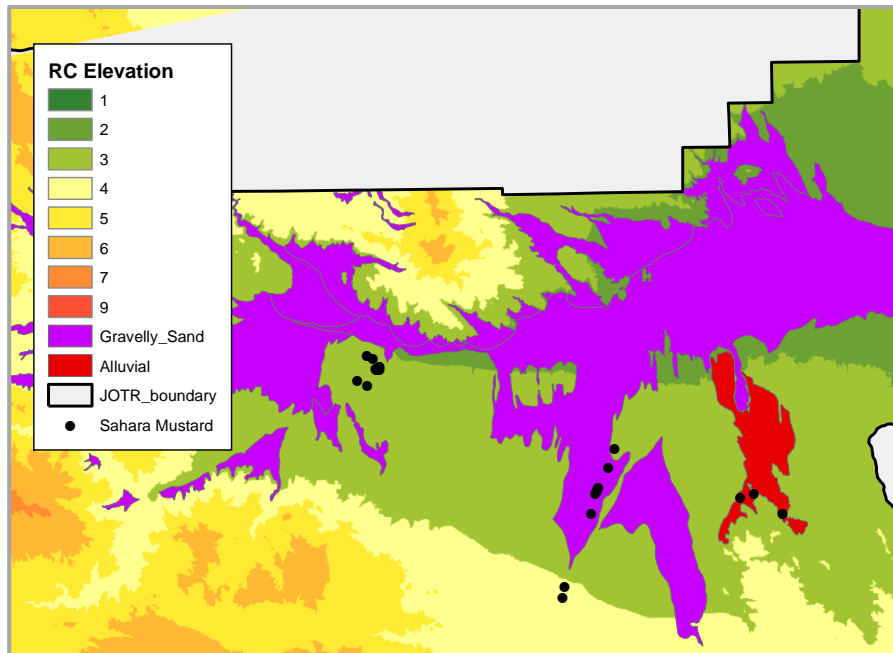
**Figure 6-3: JOTR Overview with Selected Soil and Elevation Classes**



**Figure 6-4: Identified Potential High Risk Areas for Sahara Mustard Invasion**

To test the validity of this assessment, 34 Sahara mustard weed observations points were collected in the months of February and March of 2013. The data points were imported into the geodatabase and overlaid on the assessed high risk area. The results showed only nine weed points were within the high risk area (Figure 6-5). The other weed points were located on the roadsides, a layer not used in this assessment.





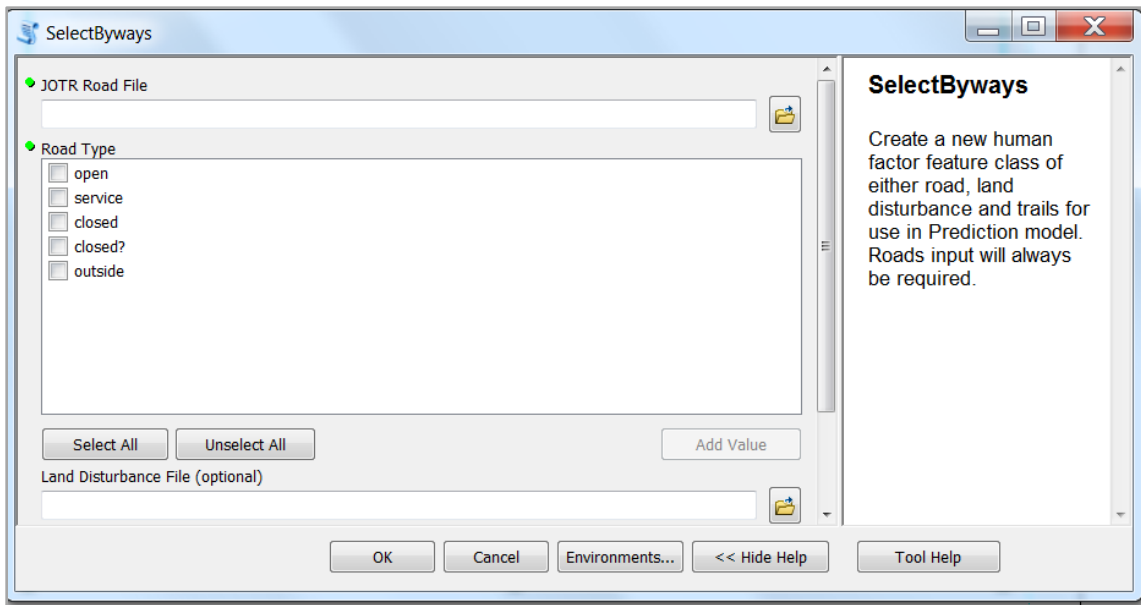
**Figure 6-5: Only Nine Mustard Weed Points in High Risk Areas**

It indicates that other environmental factors also impact the suitability, such as existing vegetation cover, aspect, slope, and climate data. These data were not incorporated into this test case due to the lack of understanding the complex relationship of all these factors and their roles in Sahara mustard weed habitats. However the developed database provides a framework for users to incorporate these factors in future habitat analysis when more information is known about the roles of various factors in Sahara Mustard habitat.

## 6.2 Results of the Predictive Spread Model

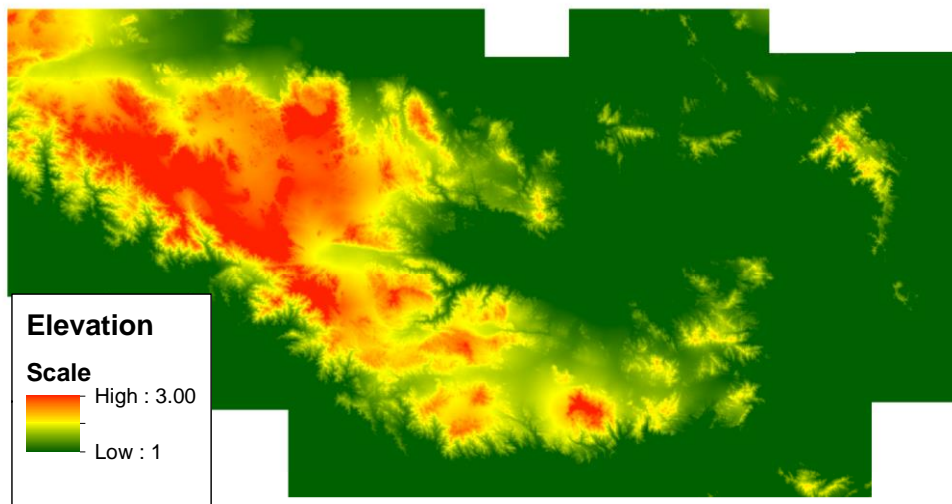
The predictive spread model forecasts potential Sahara mustard sites based on confirmed weed observations. The feature inputs for testing included the 2013 Sahara mustard data, the entire JOTR road network (which includes subtypes open, closed, and service), a rescaled elevation layer, and yearly spread rate. About 95% of Sahara mustard weed observations provided by the client were found on the roadsides of JOTR, which supports studies that concluded that the weed spreads mainly along roads. This prediction tool models this conclusion in predicting Sahara mustard spread.

Users have the option of adding the other two human factor feature layers (JOTR\_Trails, JOTR\_Land\_Disturbance) to the roads by using the SelectByways script. Figure 6-6 shows the user interface for this script.



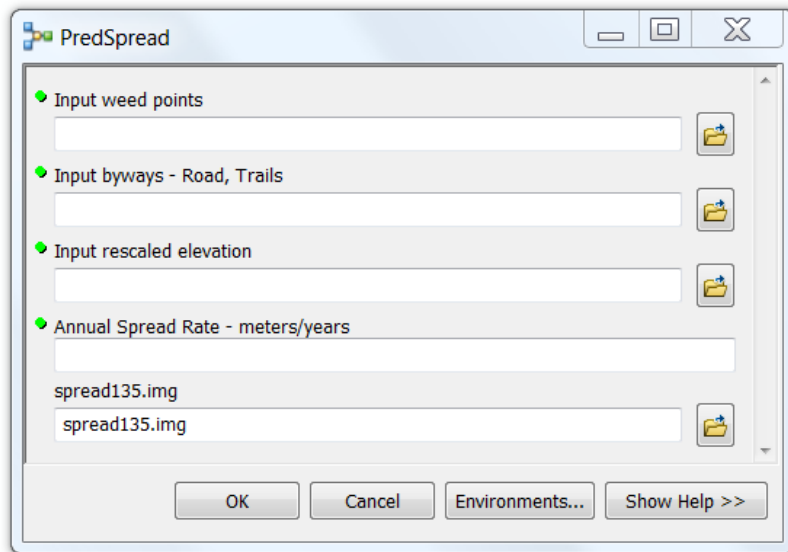
**Figure 6-6: User Interface for Script SelectByways**

The elevation layer was rescaled into a continuous range of impedance starting from 1. Figure 6-7 presents the rescaled elevation layer with values ranging from 1 to 3.



**Figure 6-7: Rescaled Elevation**

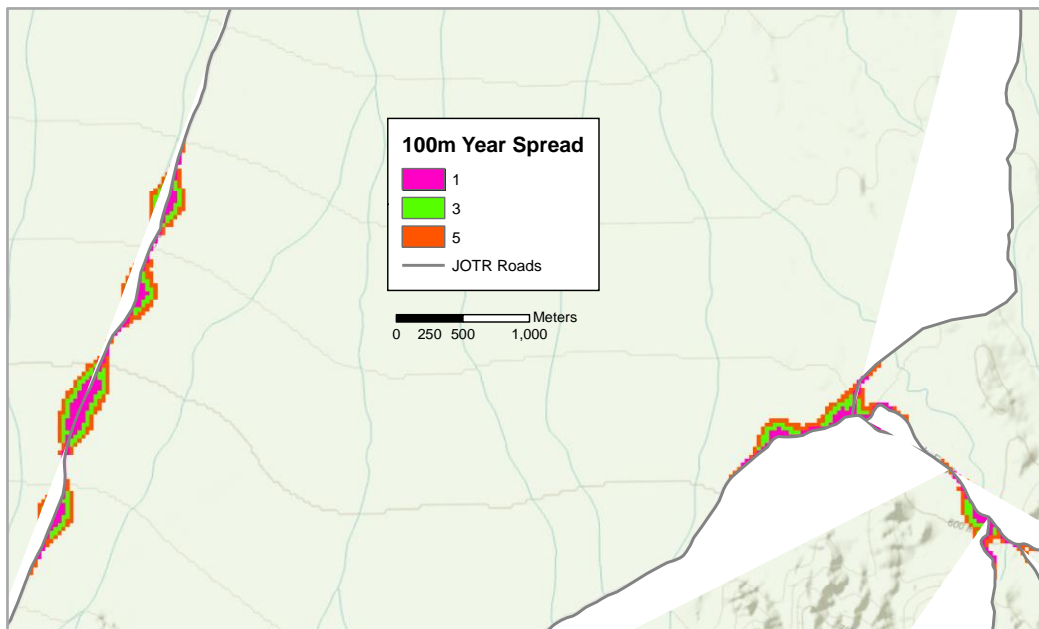
Figure 6-8 shows the user interface of the prediction tool. The process starts after all required data are provided. In this example the yearly spread rate was set for 100 meters.



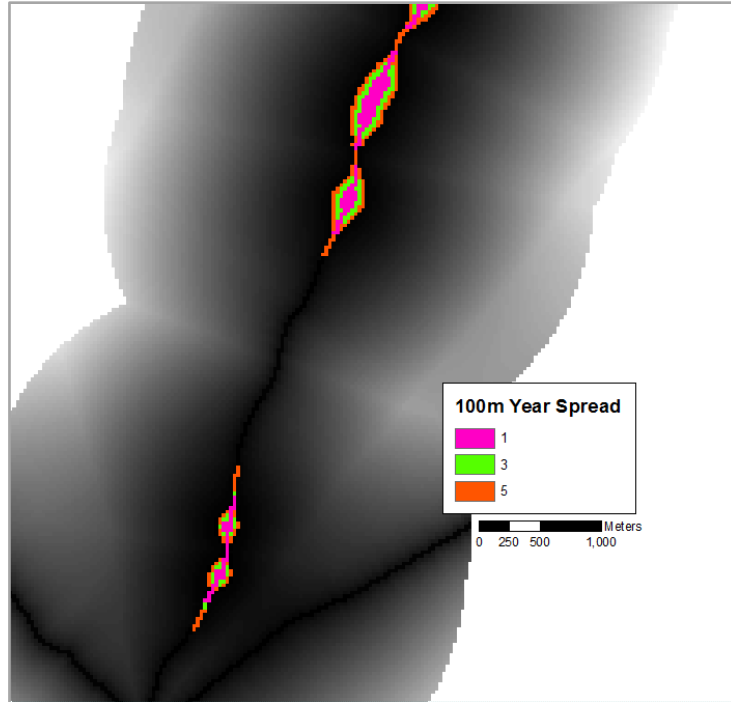
**Figure 6-8: Predictive Spread Interface**

The output of the prediction tool is a raster file, identifying the predictive spreads over periods of one year, three years, and five years. Figure 6-9 illustrates the spreads over the different periods in color.

All demonstrated spreads originate from a Sahara weed point; there are no unconnected spread cells. The impedance surface limits the spread rate off roads. Figure 6-10 shows the resulting predictive spread layer on top of the impedance surface layer.

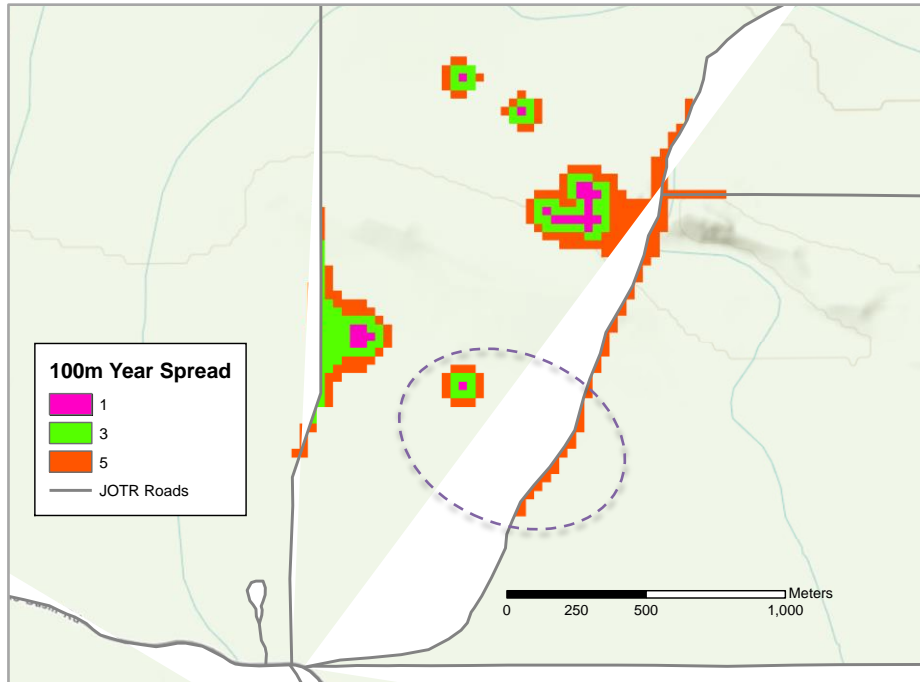


**Figure 6-9: 100 Meter per Year (1, 3, 5) Prediction Spread Results**



**Figure 6-10: Elevation Class Change Impacts the Spread Rate**

When increasing the yearly spread rate, issues can arise in certain locations with dense observed weed points. Such an issue is shown in Figure 6-11. The max spread rate is 500 meters for a five year prediction but the spread goes to over 1000 meters (indicated by the dashed circle). The programming logic for weed spread uses the conditions that the spread path be connected to a mustard weed point and be within a mustard weed point's maximum distance range. What is happening here is the program logic recognizes another nearby mustard weed point (in the dashed circle) and calculates the max spread distance from that point. The maximum spread distance from this mustard weed point is 500 meters, correct for the program logic but incorrect for the predictive spread of the originating mustard weed point. There was no check in the program logic to ensure that the conditions were based on the same weed point origin. This check should have been in the region growing process of the predictive model logic when each weed seed point's neighbors were being evaluated. The correct program logic should be for each year (1, 3, and 5) a valid predictive spread pixel is both connected to a weed seed point and within the year maximum spread distance from that seed point. This will eliminate the predictive spread going beyond the maximum spread rate.



**Figure 6-11: Example of the Spread Going Beyond the 5 Year Maximum Spread Rate**

### 6.3 Summary

This chapter presented a workflow of using the developed geodatabase in analyzing Sahara Mustard habitat. It demonstrated that other factors needed to be modeled in habitat studies. The result of the prediction tool was also reviewed.

## Chapter 7 – Conclusions and Future Work

This chapter includes a summary of this project: GIS tools for tracking the Sahara mustard weed. Recommendations for future work are then discussed.

### 7.1 Summary

The goal of this project was to develop a GIS that can provide decision support information to help the JOTR staff develop a control management plan and to prioritize eradication efforts against the Sahara mustard weed. There were two primary objectives. The first was to identify JOTR areas susceptible to weed infestation; these high risk areas will be the primary locations for the NPS to focus on when tracking down the mustard weed using only natural factors. The second objective was to provide a predictive distribution spread of the Sahara mustard weed from the influences of human factors, which are known to be strong predictors.

To meet these two objectives, a mustard weed data logical model, a geodatabase for data and land assessment analysis, and a predictive spread model were developed. The logical data model was derived from the conceptual data model that described the relationships with influencing biological (e.g., coverage, plant life stage) factors that contribute to weed spread. The result was a Sahara mustard logical data model that led to the development of the mustard weed geodatabase.

The geodatabase stores all JOTR feature layers that could be used for land assessment analysis. An analysis demonstrated that using the developed geodatabase to identify places susceptible to mustard weed invasion by overlaying specific soil types and reclassified elevation feature layers. Areas with desired soil types and within specified elevation range are deemed as high risk to weed invasion. This analysis can also help domain experts evaluate the appropriate soil types and elevation ranges in identifying vulnerable areas to weed invasion with actual weed observations.

The predictive spread model takes Sahara Mustard weed points (observations) and the JOTR human factor feature layers including roads, trails, and disturbed areas to generate potential weed spread maps for 1, 3, and 5 year periods with user provided yearly spread rates.

### 7.2 Future Work

This project provides the initial foundation for additional data and further analysis on tracking the Sahara mustard weed. Listed here are recommendations for future GIS work:

- The addition of more complete Sahara Mustard datasets (see Table 4.1) from regularly planned field Sahara mustard observations by the JOTR staff. These data will be invaluable on assessing the accurate rate of weed spread which may vary depending on the environmental and human factors that exist in the target location.
- Land cover association with the Sahara Mustard weed. Which native plants does the Sahara mustard likely be associated with? A dynamic landscape study done of the Sonoran Desert low desert produced land classification that showed the mustard weed

association mostly with desert lands occupied by small leafed desert scrub and grassland (Sanchez-Flores, 2007). This kind of field work information can help gain more knowledge about Sahara Mustard and potentially help predict occurrence of Sahara Mustard based on the existence of associated native plants.

- Weed susceptibility model based on the tally system of the Sahara Mustard's environmental habitat requirements. The more requirements a particular area has that meet the Sahara Mustard's environmental suitability the higher probability the weed can invade that location if introduced. A susceptibility layer can be developed by tallying all the environment layers and deriving the final scores that indicate susceptibility. GIS layers to include will be all soil layers, elevation and land disturbance. Other GIS layers for consideration are existing vegetation, slope, aspect and hydrology.

## Works Cited

- Andrew, M. E., & Ustin, S. L. (2009). Habitat Suitability modelling of an invasive plant with advanced remote sensing data. *Diversity and Distributions*, 15, 627-640.
- Bradley, B., Olsson, A., Wang, O., Dickson, B., Pelech, L., Sesnie, S., et al. (2012, April 2). Species detection vs. habitat suitability: Are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling*, 244, 57-64.
- Chevron Energy Solutions. (2010, Jan 14). *Weed Control Plan, The Lucern Solar Project, San Bernardino County*. San Francisco.
- Commerce, B. S. (2010). *Guidelines for treatment of Sahara Mustard Weed*. Retrieved May 3, 2013, from Sahara Mustard Weed Eradication Task Force: <http://borregospringschamber.com/mustard/documents/guidelines.htm>
- Devender, T. R. (1997). Exotic Plants in the Sonoran Desert Region, Arizona and Sonora. *California Exotic Pest Plant Council 1997 Symposium Proceedings*. Tucson: California Exotic Pest Plant Council.
- Guisan, A., & Zimmerman, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.
- Halvorson, W. L., & Guertin, P. (2003). *USGS Weeds in the West project - Factsheet for Brassica tournefortii Gouan*. Tucson: USGS Southwest Biological Science Center - University of Arizona.
- Harris, C., Parks, K. J., Atkinson, R., Edwards, C., & Travis, J. M. (2009, Sept). Invasive Species Control: Incorporating demographic data and seed dispersal into a management model for *Rhododendron ponticum*. *Ecological Informatics*, 4(4), 4, 226-233.
- Higgins, S. I., Richardson, D. M., & Cowling, R. M. (1996, Jan 20). Modeling invasive plant spread: The role of plant environment interactions and model structure. *Ecology*, 77, 2043-2054.
- Hirzel, A., Helfer, V., & Metral, F. (2001, May 22). Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, 145, 111-121.
- Johnson, W. S. (2003, 6 24). *Weed factsheet*. Retrieved October 2, 2012, from [www.unce.unr.edu](http://www.unce.unr.edu): <http://www.unce.unr.edu/publications/files/ag/2006/fs0624.pdf#search=%22sahara%20mustard%22>
- Kearney, M. (2006). Opinion: Habitat, environment and niche: what are we modelling? *OIKOS*, 115, 186-191.
- Kerr, J. T., & Ostrovsky, M. (2003, June 6). From space to species: ecological applications for remote sensing. *TRENDS in Ecology and Evolution*, 18, 299-304.
- McDonald, Chris; University of California Cooperative Extension. (2010). *Sahara Mustard Consortium*. Retrieved May 3, 2013, from Sahara Mustard Consortium: <http://saharamustard.ucanr.org/Consortium/Home.html>
- Minnich, R., & Sanders, A. (2000). *Invasive plants of California's wildlands*. Berkeley: University of California Press.
- NPS U.S. Department of the Interior. (2012). Joshua Tree Guide. *A Planning Guide for Visitors to the Joshua Tree National Park*.



- Pearlstine, L., Portier, K. M., & Smith, S. E. (2005). Textural Discrimination of an Invasive Plant, *Schinus terebinthifolius*, from Low Altitude Aerial Digital Imagery. *Photogrammetric Engineering & Remote Sensing*, 71, 289-298.
- Peterson, A. C. (1999). Sensitivity of distributional prediction algorithms to geographic data completeness tool. *Ecological Modelling*, 117, 159-164.
- Prentice, I., Cramer, W., Harrison, R. L., Monserud, R., & Soloman, A. (1992). A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, 19, 117-134.
- Renz, M. J. (2006, Mar 14). *Weed Factsheet*. Retrieved October 2, 2012, from <http://weeds.nmsu.edu>: [http://weeds.nmsu.edu/pdfs/Sahara\\_mustard\\_factsheet\\_3-14-06.pdf](http://weeds.nmsu.edu/pdfs/Sahara_mustard_factsheet_3-14-06.pdf)
- Sanchez-Flores, E. (2007). GARP modeling of natural and human factors affecting the potential distribution of the invasives *Schinus arabicus* and *Brassica tournefortii* in 'El Pinacate y Gran Desierto de Altar' Biosphere Reserve. *Ecological Modelling*, 204, 457-474.
- Taylor, C. M., & Hastings, A. (2004). Finding optimal control strategies for invasive species: a density-structure model for *Spartina alterniflora*. *Journal of Applied Ecology*, 41, 1049-1057.
- U.S. Department of the Interior. (2009). *Chevron Lucern Solar Valley Project fact sheet*. Retrieved May 5, 2013, from U.S. Department of the Interior: [http://www.doi.gov/news/pressreleases/upload/Chevron-fact-sheet\\_1004.pdf](http://www.doi.gov/news/pressreleases/upload/Chevron-fact-sheet_1004.pdf)
- USDA Forest Service Remote Sensing Applications Center. (2011). *A Weed Manager's Guide to Remote Sensing and GIS*. Retrieved 10 2012, from UDSA FS Invasives Species: <http://www.fs.fed.us/eng/rsac/invasivespecies/index.htm>
- USGS Western Ecological Research Center. (2009). *Invasive weed removal and habitat restoration: Sahara Mustard*. Retrieved Oct 3, 2012, from USGS WERC: <http://www.werc.usgs.gov/Project.aspx?ProjectID=95>
- Williams, D. G., & Baruch, Z. (2000). African Grass Invasion in the Americas: Ecosystem Consequences and the role of Ecophysiology. *Biological Invasions*, 2, 123-140.

## Appendix A. Python String Function

```
def soilSQL(listSoil):
    """Take list and convert into SQL string"""
    soilString = "("
    for x in listSoil:
        if x == 'sand':
            soilString = soilString + "sand',"
        elif x == 'fine sand':
            soilString = soilString + "fine sand',"
        elif x == 'fine sandy loam':
            soilString = soilString + "fine sandy loam',"
        elif x == 'loamy sand':
            soilString = soilString + "loamy sand',"
        elif x == 'gravelly sand':
            soilString = soilString + "gravelly sand',"
        elif x == 'gravel loamy sand':
            soilString = soilString + "gravel loamy sand',"
        elif x == 'very gravelly loam':
            soilString = soilString + "very gravelly loam',"
        elif x == 'very gravelly sandy loam':
            soilString = soilString + "very gravelly sandy loam',"
        elif x == 'extremely gravelly loam':
            soilString = soilString + "extremely gravelly loam',"
        elif x == 'extremely gravelly sandy loam':
            soilString = soilString + "extremely gravelly sandy loam',"
        elif x == 'rock outcrop':
            soilString = soilString + "rock outcrop',"
        elif x == 'rock outcrop association':
            soilString = soilString + "rock outcrop association',"
        elif x == 'rock outcrop complex':
            soilString = soilString + "rock outcrop complex',"
        elif x == 'association':
            soilString = soilString + "association',"
        elif x == 'complex':
            soilString = soilString + "complex',"
        else:
            soilString = soilString + x + ","

    soilLastChar = len(soilString)
    soilString = soilString[0:soilLastChar-1]
    soilString = soilString + ")"

    return soilString
```

## Appendix B. Rescale Tool Code

Algorithm for rescale method:

If Elevation = threshold value then

    Rescale\_value = 1

Else

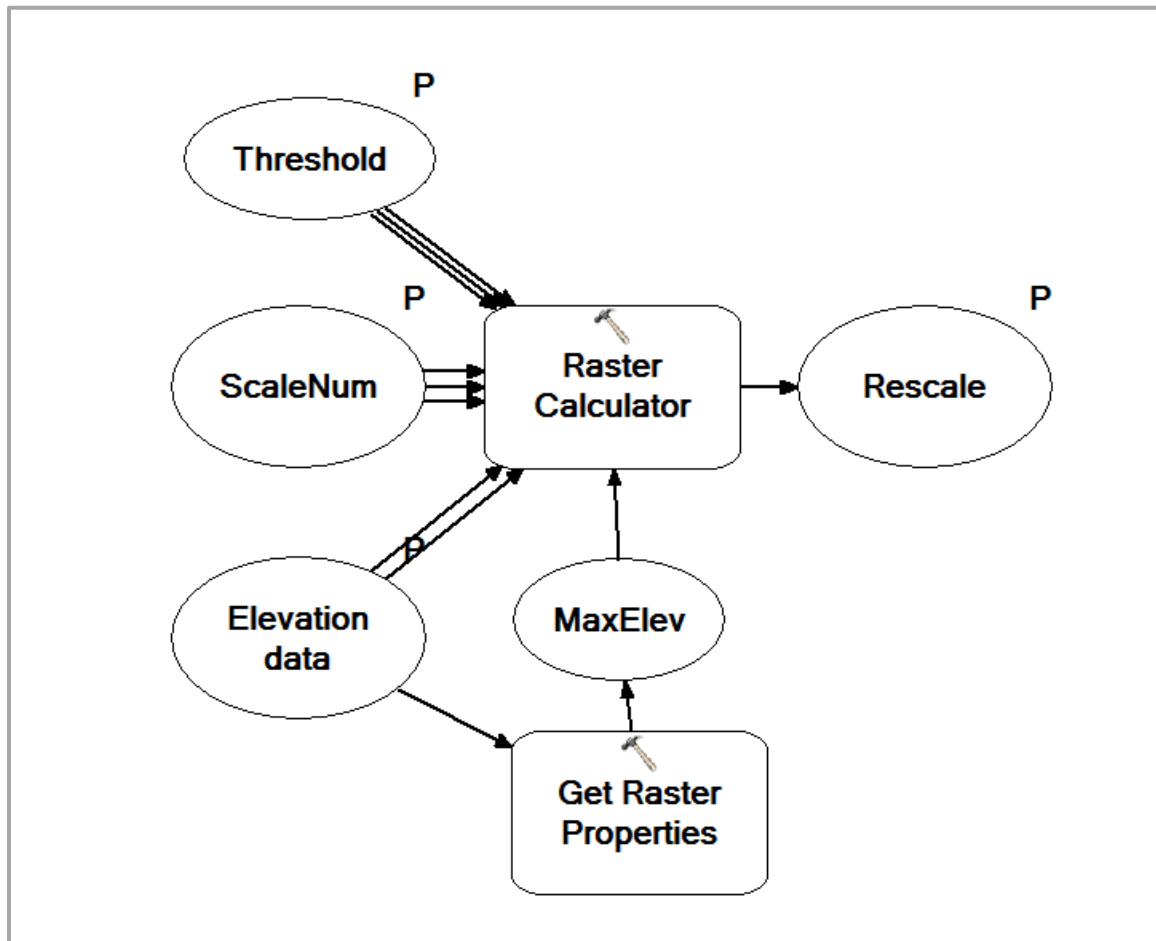
    Rescale\_value =

$1 + (\text{elevation} - \text{threshold}) * (\text{num of Scales} - 1) / (\text{max elevation} - \text{threshold})$

Transformed into Raster Calculator format for ModelBuilder:

```
Con("%Elevation data%" <= (%Threshold%),1,(1+("%Elevation data%" - (%Threshold%)) * ((float(%ScaleNum%))-1) / (float(%MaxElev%) - (%Threshold%))))
```

ArcPy's Get Raster Properties is used to extract the maximum elevation from the input elevation file. The figure below demonstrates the ModelBuilder data flow.



## Appendix C. Region Growing Code

```
# Import system modules
import arcpy
from arcpy import env
from arcpy.sa import *
import numpy
from osgeo import gdal, gdalconst
import sys
import math

arcpy.env.overwriteOutput = True

##PARAMETERS - INPUTS BY USER #####
cost_layer = arcpy.GetParameterAsText(0)
euc_weed = arcpy.GetParameterAsText(1)
#spread rate for the year
spread_rate_in = arcpy.GetParameterAsText(2)
out_spread = arcpy.GetParameterAsText(3)
spread_rate = int(spread_rate_in)
#input raster data transform into Numpy arrays for processing
cost_array = arcpy.RasterToNumPyArray(cost_layer)
weed_array = arcpy.RasterToNumPyArray(euc_weed)
rasterIn=gdal.Open(cost_layer, gdalconst.GA_ReadOnly)
#####

#Retrieve file format information for use in processing and out file
colNum=rasterIn.RasterXSize #Number of columns
rowNum=rasterIn.RasterYSize #Number of rows
projectionfrom = rasterIn.GetProjection()
geotransform = rasterIn.GetGeoTransform()

#create working and output rasters, initiate with all zeros
imgSeg=numpy.zeros((1,rowNum,colNum),dtype=numpy.float32)
imgSeg1=numpy.zeros((1,rowNum,colNum),dtype=numpy.float32)
imgSeg3=numpy.zeros((1,rowNum,colNum),dtype=numpy.float32)
imgSeg5=numpy.zeros((1,rowNum,colNum),dtype=numpy.float32)
outRaster=numpy.zeros((1,rowNum,colNum),dtype=numpy.float32)

#list to hold seeds (weed points) for capturing only neighboring
#cells that are within the threshold values
seedYRList = []

#####
##PROCESS 1 - capture only cells that are within both spread rate distance
## and threshold value - store in imgSeg (year 3) and imgSeg1 (year 1)
```

```

for r in range( rowNum ) :
    for c in range( colNum ) :

        if (weed_array[r,c] >= 0 and weed_array [r,c]< (spread_rate * 5)):
            if cost_array [r,c]< 1250 :
                #outRaster[0,r,c]= 5
                imgSeg5[0,r,c] = 5
            if (weed_array[r,c] >= 0 and weed_array [r,c]< (spread_rate * 3)):
                if (weed_array[r,c] == 0): #create seedlist from mustard weed points
                    newSeed = [r,c]
                    seedYRList.append(newSeed)
                if cost_array [r,c]< 750:
                    imgSeg3[0,r,c] = 3
            if (weed_array[r,c] >= 0 and weed_array [r,c]< spread_rate):
                if cost_array [r,c]< 250:
                    imgSeg1[0,r,c] = 1
##end for loop

#####
##PROCESS 2
##Region growing method -Remove from captured cells both imgSeg[], cells that
##are not directly connected to a seed in seedYRList[]
imgSeg[0,:,:] = imgSeg5[0,:,:]

for year in range (5,0,-2):
    featureList = []
    seedList = []
    seedmasterList = seedYRList

    while seedmasterList != []:
        for seed in seedmasterList:
            featureList.append(seed)
            #eval neighbor are within the min max range of the image's row & col
            A = 0 #row - 1
            B = 0 #row + 1
            C = 0 #col - 1
            D = 0 #col + 1
            if ((seed[0]-1) >= 0):          #seed[0],seed[1] = row,col
                A = 1
            if ((seed[0]+1) <= rowNum-1):
                B = 1
            if ((seed[1]-1) >= 0):
                C = 1
            if ((seed[1]+1) <= colNum-1):
                D = 1
            #eval seed neighbors to be new seeds if selected put into new seed list

```

```

#and turn into 0 back in image
if A:
    if imgSeg[0,seed[0]-1,seed[1]] == year:
        newSeed = [seed[0]-1,seed[1]]
        seedList.append(newSeed)
        imgSeg[0,seed[0]-1,seed[1]] = 0
    if C:
        if imgSeg[0,seed[0]-1,seed[1]-1] == year:
            newSeed = [seed[0]-1,seed[1]-1]
            seedList.append(newSeed)
            imgSeg[0,seed[0]-1,seed[1]-1] = 0
if B:
    if imgSeg[0,seed[0]+1,seed[1]] == year:
        newSeed = [seed[0]+1,seed[1]]
        seedList.append(newSeed)
        imgSeg[0,seed[0]+1,seed[1]] = 0
    if D:
        if imgSeg[0,seed[0]+1,seed[1]+1] == year:
            newSeed = [seed[0]+1,seed[1]+1]
            seedList.append(newSeed)
            imgSeg[0,seed[0]+1,seed[1]+1] = 0
if C:
    if imgSeg[0,seed[0],seed[1]-1] == year:
        newSeed = [seed[0],seed[1]-1]
        seedList.append(newSeed)
        imgSeg[0,seed[0],seed[1]-1] = 0
    if B:
        if imgSeg[0,seed[0]+1,seed[1]-1] == year:
            newSeed = [seed[0]+1,seed[1]-1]
            seedList.append(newSeed)
            imgSeg[0,seed[0]+1,seed[1]-1] = 0
if D:
    if imgSeg[0,seed[0],seed[1]+1] == year:
        newSeed = [seed[0],seed[1]+1]
        seedList.append(newSeed)
        imgSeg[0,seed[0],seed[1]+1] = 0
    if A:
        if imgSeg[0,seed[0]-1,seed[1]+1] == year:
            newSeed = [seed[0]-1,seed[1]+1]
            seedList.append(newSeed)
            imgSeg[0,seed[0]-1,seed[1]+1] = 0

#clear all processed seeds from seedmasterList, append new seed list
#and process starts all over with new seed list
seedmasterList = []
for seed in seedList:

```

```

        seedmasterList.append(seed)
    #clear out new seedList
    seedList = []
##end of While loop #####

#Output the featureList which are the cells who are connected to a seed
for pixel in featureList:
    outRaster[0,pixel[0],pixel[1]] = year
    if year == 5:
        imgSeg[0,:,:] = imgSeg3[0,:,:]
    if year == 3:
        imgSeg[0,:,:] = imgSeg1[0,:,:]
##end for
#####
####

#Output the resulting 1,3 & 5 year spread to raster file
rasterOut1=gdal.GetDriverByName('HFA').Create(out_spread,colNum, rowNum,1,
gdalconst.GDT_Float32)
rasterOut1.SetProjection(projectionfrom)
rasterOut1.SetGeoTransform(geotransform)
rasterOut1.GetRasterBand(1).WriteArray(outRaster[0,:,:])

rasterOut1=None
rasterIn=None

print "Script Complete"

```