Predictive Modeling of Sulfur Flower Buckwheat (Erigonum umbellatum Torrey) Using Non-Parametric Multiplicative Regression Analysis

David B. Davis
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Animal Sciences Commons

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/etd/1940

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Predictive Modeling of Sulfur Flower Buckwheat (*Eriogonum umbellatum* Torrey) Using Non-Parametric Multiplicative Regression Analysis

D. Bracken Davis

A thesis submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Master of Science

Steven L. Petersen
Bruce A. Roundy
Mark W. Jackson

Department of Plant and Wildlife Sciences
Brigham Young University

December 2009

Copyright © 2009 D. Bracken Davis

All Rights Reserved
ABSTRACT

Predictive Modeling of Sulfur Flower Buckwheat (*Eriogonum umbellatum* Torrey) Using Non-Parametric Multiplicative Regression Analysis

D. Bracken Davis

Department of Plant and Wildlife Sciences

Master of Science

Impacts of humans on ecosystems in western United States have necessitated ecological restoration, which includes the development of native seed that can be used for revegetation efforts. Development of such seed sources are costly and time consuming. This study describes the use of non-parametric multiplicative regression analysis (NPMR) to develop a predictive model for occurrence of sulfur-flower buckwheat (*Eriogonum umbellatum* Torrey) population seed collection. This perennial forb species is of interest for seed source development in the western United States. Presence and absence data for *E. umbellatum* was taken from the Utah Division of Wildlife Resources Big Game Range Trend project as well as herbarium specimens across Utah, U.S.A. NPMR, a statistical niche modeling system that selects the best predictor variables and develops probability of occurrence estimates multiplicatively, was used to select predictor variables from spatially explicit data made available in a Geographic Information System (GIS). Two models were created using NPMR, one with a suggested default minimum average neighborhood size and the other with a less-restricted minimum average neighborhood size. GIS maps of models were created, artificially classified into low, medium, and high probability areas, and validated in the field in Tooele County, Utah. Of 68 possible physiographic, climatic, and soil variables provided for analysis, NPMR selected 4 variables for the default minimum average neighborhood model and 10 variables for the less restricted neighborhood model. The default model had a higher descriptive statistic (log β value) and mapped a larger area than the less restrictive neighborhood model. When increased minimum neighborhood sizes were selected during the development of the probability maps, the resulting areas of probability prediction decreased. The presence rates of *E. umbellatum* in field-validated test sites were 7.4%, 12.0%, and 28.6% for the low, medium, and high probability sites, respectively. Although presence rates of field validated data were lower than the predicted probability ranges for those same sites, presence rates increased with increased probability ranges. Using the generated model can reduce the cost and time necessary to locate plants compared to searching for species populations using an undirected approach.

Keywords: Probability mapping, niche theory, Great Basin, GIS, seed source development.
ACKNOWLEDGEMENTS

I appreciate Steve Petersen, Bruce Roundy, and Mark Jackson for their consultation and critical review of the project and this manuscript. I would like to thank the Utah Division of Wildlife Resources for providing the extensive dataset used for this study as well as years of training on the collection and use of that data. I appreciate the Range Trend summer field crews who labored diligently for the collection of this data. Finally, I thank Andrew Yost for his suggestions when getting the project started. Carmen Davis provided support, advice, and invaluable editorial contributions.
Brigham Young University

SIGNATURE PAGE

of a thesis submitted by

D. Bracken Davis

The thesis of D. Bracken Davis is acceptable in its final form including (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory and ready for submission.

Date ________________________________________________

Steven L. Petersen, Chair

Date ________________________________________________

Bruce A. Roundy, Committee Member

Date ________________________________________________

Mark W. Jackson, Committee Member

Date ________________________________________________

Loreen Allphin, Graduate Coordinator

Date ________________________________________________

Rodney J. Brown, College of Life Sciences Dean
# TABLE OF CONTENTS

Abstract ................................................................................................................................. ii

Acknowledgements ............................................................................................................... iii

Table of Contents ................................................................................................................ v

List of Figures ....................................................................................................................... vi

Introduction .......................................................................................................................... 1

Methods ............................................................................................................................... 3

  Study Area ......................................................................................................................... 4

  Input Data (Response Variables) ....................................................................................... 5

  Predictor Variables ........................................................................................................... 6

  Data Preparation ............................................................................................................. 10

  Analysis ............................................................................................................................ 11

  Model Testing .................................................................................................................. 12

Results .................................................................................................................................. 13

Discussion ............................................................................................................................ 15

Literature Cited .................................................................................................................... 20
LIST OF FIGURES

Fig. 1. Relief map and location of Utah in the United States. The area within the white border is Tooele County. ................................................................. 27

Fig. 2. Map of the SSURGO soil data coverage for the state of Utah. The solid gray areas are those with soil data coverage. ........................................................................................................... 28

Fig. 3. Comparison of the subset predictive maps with map creation input minimum neighborhood sizes of 1, 3, and 10. White areas are those with no predicted species occurrence, light gray are those with low probabilities, those with dark gray are medium probabilities, and black are those with high probabilities. The input minimum neighborhood size of the map on the left is 1, the center is 3, and right is 10. As the minimum neighborhood size increased, the area displayed on the predictive map decreased. ........................................................................ 29

Fig. 4. Predictive map of *Eriogonum umbellatum* within the study area. The predictor variables selected during the analysis of this model included elevation, slope, land cover, and soil saturated hydraulic conductance (ksat). The areas with no predicted species occurrence are displayed in white, those with low probability in light gray, those with medium probability in dark gray, and those with high probability in black. ............................................................................................... 30

Fig. 5. Distribution of field verification sample sites across the Tooele County predictive map................................................................. 31

Fig. 6. Relief map of the study area with field verification sample sites and watershed boundary. The red points are those with expected low probability of occurrence, those with yellow are medium, and those with green are high. Circular points are sites with species absence and star points are species presence. The blue line represents the watershed boundary for the watershed..................................................................... 32
INTRODUCTION

Human impacts on the western United States have necessitated that natural resource managers increase ecological restoration efforts across the landscape. These impacts include urbanization (Mills et al. 1989, Frissell 1993), exploitation of natural resources (Frissell 1993), poor (or ineffective) management strategies (Keeley 2006), and the introduction of exotic species (Pimental et al. 2000, Thompson et al. 2006). Many exotic species replace native species and permanently change ecosystem function (D’Antonio and Chambers 2006). It is therefore imperative to identify means of maintaining sustainable ecosystem function through development of restoration methods that minimize human impact. Historically, introduced species have been applied in restoration work because of high success and adaptation to compete with exotic weeds (Thompson 2006), however introduced species are not necessarily ideal for restoring native landscapes due to their ability to outcompete native species (Monsen 2004, Monsen and Stevens 2004, Stevens 2004, Waldron et al. 2005). Substantial resources by state, federal, and research agencies are being allocated into developing seed sources of native plants which successfully establish and compete with exotics and maintain ecosystem function.

One species of interest for seed source development in Utah is sulfur flower buckwheat *Eriogonum umbellatum* Torrey (Polygonaceae). Several characteristics of *E. umbellatum* make it a desirable species for seed development. It is a ubiquitous species found in many habitats within Utah and seed collection is relatively simple for seed source development. It is also believed to be an important species for wildlife species in Utah (Jason Vernon, personal communication, 2007). Shannon et al. (1975) specifically showed that *E. umbellatum* is an important Rocky Mountain Bighorn Sheep (*Ovis canadensis*) summer range forage species.
*Eriogonum umbellatum* is a low-growing, often mat-forming, perennial forb distributed across the western United States. Creamy white to yellow flowers suffused with red or purple are born on umbellate inflorescences that often grow above the vegetative structure of the plant making seed harvesting simple. Welsh et al. (1993) describe four varieties in Utah, however, Flora of North America (1993+) describes eight varieties in Utah, thus indicating some ambiguity in taxonomic classification. These varieties vary in habitats from Wyoming big sagebrush (*Artemesia tridentata ssp. wyomingesis* Beedle & A. Young) to alpine meadow and mountain talus communities. Several varieties have been used ethobotanically by indigenous people of North America (Flora of North America, 1993+).

Due to its wide distribution throughout Utah, high abundance in certain populations, ease of seed harvest, and importance for wildlife, *E. umbellatum* is a good candidate for seed development. However, the financial costs and time requirements of locating potential seed source populations by simply surveying areas with suspected population distributions are extremely high. This necessitates a low-cost modeling approach for locating and mapping potential seed source populations. Scientists have examined varied methods of modeling to predict species distributions based on ecological niche theory (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith et al. 2006, Kearney and Porter 2009). Coupled with Geographic Information System (GIS) tools, the distribution of mapped environmental input variables that correlate with known species distributions can then be applied to develop probability of occurrence models from which predictive distribution maps are produced.

In this paper, we investigate the feasibility of predicting *E. umbellatum* distribution using non-parametric multiplicative regression analysis (NPMR), a method recently developed by McCune (2006). Yost (2008) explains that NPMR analyzes environmental gradients, or
predictor variables, against locations with known observations of the species of interest by using kernel functions to weigh those observations multiplicatively, rather than using an additive approach typical of many models. The interactions of environmental variables in nature are complex and often are not additive, therefore a multiplicative approach may potentially model those natural interactions better than the additive approach. McCune (2006) gives examples of the difficulty of using additive models to describe complex ecosystem variable interactions. The majority of models require that a priori assumptions about the species response to environmental variables be made (Austin 2002). However, the approach used in NPMR uses a multiplicative kernel smoother method to analyze the effect of each predictor variable on species distribution and is based on the interactions between predictor variables (McCune 2006). NPMR has been used to successfully predict the potential responses of species following different climate change scenarios (Ellis et al. 2007), distribution of introduced species in estuaries along the western coast of the United States (Ruesser and Lee 2008), and the distribution of indicator plant species in forests of northwestern United States (Yost 2008). The purpose of this study was to develop a predictive distribution map of *E. umbellatum* in western Utah for seed source populations using NPMR and provide land managers with a map to predict the potential success of seeding *E. umbellatum* at a given location.

**METHODS**

For this study, we followed a standard species habitat modeling procedure, where: 1) study field data is collected; 2) predictor variable data are compiled and prepared for analysis; 3) field data and predictor variables are analyzed; 4) alternate models are created; 5) models are
tested in the field; 6) model accuracy is determined from field results; and, 7) models are possibly refined and further accuracy assessments performed (Franklin 1995, Yost 2008).

Two types of data were required for the analysis and predictive modeling phases: 1) response variables, which included location data (GIS) and presence or absence of the species being modeled; and 2) the predictor or environmental variables, from which input site characteristics are taken and are later used as the reference for the spatial predictive interpolation.

Study Area

The input response variable data for the model was taken from data sources that spanned across the State of Utah. Utah is comprised of the following six different ecoregions: the Great Basin along the western half of the state, the Colorado Basin along the eastern half, the Wasatch Mountain Range down the center, the Wyoming Basin along the northeast borders, the Northern Basin and Range in the northwest corner, and the Mojave Desert in the southwest corner. Average annual precipitation in Utah ranges from 127-1,854 mm (U.S. Department of Agriculture-Natural Resources Conservation Services and Oregon State University 1999) and ranges in elevation from 664-4,948 m, spanning many ecosystems.

Early in the study, it was recognized that the extent of the state was far too great to compute the model because of modeling software restrictions and computation time. Therefore, the extent of the study was restricted to the County of Tooele, along the western border of Utah within the Great Basin and the Lake Bonneville Basin (Fig. 1). This county was selected because it is mainly public land, is extensive and accessible, has high diversity in land forms and vegetation community types, and is relatively characteristic of other locations within the Great
Basin. The county contains three large military installations which were excluded from the study.

Tooele County’s average annual precipitation ranges from 127-1,245 mm (U.S. Department of Agriculture-Natural Resources Conservation Services and Oregon State University 1999) and the elevation ranges from 1,279-3,355 m, with lower precipitation in low dry desert flats to higher precipitation in high desert mountain ranges.

Input Data (Response Variables)

The response variable data, or data which shows the presence or absence of a particular species, was taken from the following four sources: Utah Division of Wildlife Resources Big Game Range Trend Studies (UDWRRT), herbarium specimens, US Forest Service (USFS) Shrub Sciences Laboratory plant materials research, and personal field observations. The bulk of the data source was provided by the UDWRRT (Summers et al. 2007), a project which has been monitoring wildlife and habitat changes in Utah since 1983. The UDWRRT project collects species cover and nested frequency (abundance) data on a five-year rotational basis. The UDWRRT sampling methods include modified Daubenmire cover class and nested frequency estimations (Daubenmire 1959, Summers et al. 2007). These estimations are taken on 100, 0.25 m² quadrats measured along five 152.4-m transects. These transects are randomly placed perpendicularly along baseline transects that range from 61.0 to 152.4 m in length. The GPS coordinate location for each UDWRRT sample site was taken at the beginning of every baseline transect. For each UDWRRT study location, if *E. umbellatum* had occurred since 1983, the study was given a presence designation. Although the presence and absence data for 909
UDWRRT studies were originally compiled for the study, only 644 studies were included because the other 265 were in areas missing one or more predictor variables.

Another 36 locations with presence/absence data were taken from herbarium specimens from Utah Valley University, University of Utah, Idaho State University, and University of Nevada Las Vegas. Only specimens with GPS locations were selected for the study. Sixty-six locations were provided by the USFS plant materials research team and 53 from personal observations while traveling the state before the study was conducted. The combined total of locations with *E. umbellatum* presence/absence data was 799, 139 of which had *E. umbellatum* present.

**Predictor Variables**

Predictor variables tested in the analysis and used to build predictive maps can be separated into two groups: 1) those which come from individual data-specific sources and describe soil or plant community characteristics; and 2) those which estimate the physiographic characteristics of the landscape, all derived from digital elevation data. Both data groups are continuous raster datasets that were used for both analysis and GIS predictive model construction. All raster data were prepared for analysis with the ArcGIS 9.3 (Environmental Systems Research Institute) software package. A total of 68 predictor variables were prepared for analysis of Utah and predictive mapping in Tooele County.

The bulk of the predictor variables (44 variables) were soil characteristics derived from the United States Department of Agriculture Natural Resources Conservation Service Soil Survey Geographic (SSURGO; Natural Resources Conservation Service, United States Department of Agriculture 2008) data. The SSURGO data are collected and digitized in survey
areas that range in size and availability, therefore all available survey area raster datasets were created from the database and merged together. The soils data covered approximately 67% of the state at the time of the study (Fig. 2). Qualitative soil variables included: Drainage class—all components, drainage class—dominant component, ecological site name, ecological site identification number, soil map unit name, parent material name—all components, parent material name—dominant component, soil texture—all components, soil texture—dominant component, soil texture—all components with only the basic textures coded (18 possible codes), and soil texture—all components with all possible texture codes (142 possible codes, which included the 18 codes such as gravel, bedrock, etc).

Some quantitative soil variables could be derived for user-defined soil depths. To identify differences in rooting depth needs of *E. umbellatum* in the analysis, variables were created with rooting depths of: surface, 0-15cm, 15-25cm, 25-50cm, and 50-100cm. Soil characteristics categorized into these depths include: available water capacity, calcium carbonate concentration, cation-exchange capacity, electrical conductivity, percent organic matter, and pH. Saturated hydraulic conductivity (ksat), soil depth to a restrictive layer, and average water table depth for the entire calendar year were not categorized by depth because the data was collected for the entire soil profile.

Site northing and easting were calculated for each site location using the Calculate Geometry function in the ArcGIS 9.3 software package. Land cover and land form raster datasets from the Southwest Regional GAP analysis (SWReGAP; Lowry et al. 2005) were also included. The SWReGAP data was anticipated to indicate if broad-scale plant community and land feature data were useful in creating a predictive model.
Climate data was suspected as an important predictor for *E. umbellatum* distribution, but little fine-scale raster climate data has been created. However, 1-km resolution Parameter-Elevation Regressions on Independent Slopes Model (PRISM; PRISM Climate Group, Oregon State University 2008) data for average annual maximum temperature, annual minimum temperature, annual precipitation, and monthly precipitation from the period of 1961 to 1990 was resampled to a 10-m resolution and used in the analysis.

National Elevation Dataset (NED) 1/3 Arc Second (10 m) elevation data were acquired from the USGS Seamless Server for the state of Utah. All NED datasets were merged together and filled, to remove pits and peaks in the dataset, using ArcGIS 9.3 software (ESRI). Raster datasets of degrees slope and degrees aspect were derived from the NED data using the Spatial Analyst tools in ArcGIS 9.3. The aspect dataset was then categorized as north, south, east, west, northwest, southwest, northeast, and southeast.

A heat load index (HLI) dataset was created based on a method developed and field tested by McCune and Keon (2002), which was later corrected and refined by McCune (2007). This index estimates the relative amount of heat load from incident radiation for a given location. The method proposes three possible equations to derive an HLI from slope, aspect, and latitude. The three equations can be selected for analysis based on the study location and range of slopes. Equation 1 has the broadest application, but has the lowest $R^2$ (0.958), and can be applied to slopes $\leq 90^\circ$ in steepness and between latitudes of 0-60° N. Equation 2 has a higher $R^2$ (0.978), but can only be applied to slopes with the steepness $\leq 60^\circ$. Equation 3, with yet a slightly higher $R^2$ (0.983), but can only be applied to slopes with the steepness $\leq 60^\circ$ and between the latitudes of 30-60° N. McCune and Keon (2002) also provided an equation for “folding” the aspect such
that northeast has the lowest heat load and southwest has the highest:

\[
\text{Folded aspect} = | \pi - \text{aspect} - (5\pi/4) |
\]

The HLI equations require that the slope and aspect data be in radians, therefore these raster data sets were converted to radians in the ArcGIS 9.3 Raster Calculator. We then entered the aspect in radians raster dataset into the above equation in ArcGIS 9.3 to calculate the folded aspect in radians. All study plots were \( \leq 60^\circ \) slope steepness and the latitude of the approximate center of Utah was 39.38\(^\circ\), therefore we determined it best to use Equation 3:

\[
\text{Equation 3} = 0.339 + [0.808 \times \cos(l) \times \cos(s)] - [0.196 \times \sin(l) \times \sin(s)] - [0.482 \times \cos(a) \times \sin(s)]
\]

where \( l \) is the latitude of the center of the study in radians, \( s \) is the slope in radians, and \( a \) is the folded aspect in radians. The result is a value from 0 to 1, where 0 is the lowest heat load and 1 is the highest. This equation was calculated in ArcGIS 9.3 using the slope and folded aspect raster datasets.

Iverson et al. (1997) developed an Integrated Moisture Index (IMI), which is a method for estimating relative moisture availability of a given landscape based on topography. The three topographic features, all of which were derived from the NED data in the ArcGIS 9.3 spatial analyst tools, include hillshade, flow accumulation, and curvature. The hillshade data was computed with the hillshade tool, the flow accumulation with the flow direction and flow accumulation tools, and the curvature with the curvature tool. In ArcGIS 9.3, each of the three
data types were then reclassified and normalized on a scale of 0-100. The curvature dataset was
inversed to provide higher values for basins of water collection and lower values for hilltops. In
ArcGIS 9.3, the following equation was used to add weight to and combine the three topographic
data types together for the IMI value:

\[
IMI = (\text{hillshade} \times 0.5) + (\text{curvature} \times 0.15) + (\text{flow accumulation} \times 0.35)
\]

The index output is on a scale of 0 to 100, where 0 represents no moisture
accumulation/retention and 100 represents the highest moisture accumulation and retention.

Data Preparation

The GPS coordinate locations for all response variable locations were combined into a
single GIS layer. Around each point, 20 m and 100 m buffers were created to roughly represent
the sample areas of the response variables. The Zonal Statistics function in ArcGIS 9.3 was used
to obtain predictor variable statistics for each response variable location. The 100 m buffers
were used to capture the value that made up the majority of raster cells for categorical predictor
and soil data sets. Because of the heterogeneous nature of some of quantitative predictor
variables, the smaller 20 m buffer was used to calculate the mean raster value of those data sets.

The statistics for the predictors were compiled into a single data set and response variable
sites that did not have soils data were excluded from the study to remove bias toward non-soil
datasets in analysis.
Analysis

NPMR parsimoniously models the probability of occurrence for a given species, or suite of species, by analyzing multiple ecological variables multiplicatively. McCune (2006) and Yost (2008) describe the process of model selection using a kernel weighting function to select predictor variables based on their relationship to species occurrence or abundance. There are two phases in the process of NPMR analysis: 1) the calibration phase, which is used to select the best set of predictor variables, determine a standard deviation (tolerance) value for the continuous predictor variables, and to rank each model based on a descriptive statistic; and 2) the application phase, which is used to predict species occurrence or abundance based on the predictors selected in the calibration phase.

Model strength in NPMR is determined by the descriptive statistic $\log \beta$, which is the log likelihood of ratios for two competing models (McCune 2006 and Yost 2008). The $\log \beta$ is sensitive to the number of response variables and therefore can become larger with a larger sample size. It begins at 0 and as the predictive strength of a model increases, so does the $\log \beta$ value. Yost (2008) indicates that a $\log \beta$ ranging from 1.0-2.0 is strong and $>2.0$ is decisive.

We used the Local Mean – Gaussian (NPMR) modeling function in Hyperniche 1 software (McCune and Medford 2004) for all analysis and to fit models to the predictor and response variables. This function uses a step-wise free search to identify the predictor combinations with the highest $\log \beta$ value when compared with the response variables. A variable that increased the $\log \beta$ value of the yet largest model by at least 5% was retained and selected as the best model. In this study, models with an average neighborhood size of forty (or 5% of the sample size) or less were excluded from analysis, which is the Hyperniche default
setting. After the best model was selected from all variable combinations, that model was “fine tuned,” which adjusted the continuous variable tolerances by 1% of the variable’s value range (Yost 2008).

To identify whether the default minimum average neighborhood size setting of 5% of the sample size was appropriate for the study, models with neighborhood sizes of 40 (5% of the sample size and the default) and 100 (12.5% of the sample size) were created and compared. The better of these two models, based on log $\beta$ values, was selected for further testing.

Predictive estimation maps were created using the GIS function in Hyperniche for the models with the highest log $\beta$ values selected in the calibration phase. A subset of Tooele County was selected to test the minimum neighborhood size for estimation map creation. This was performed to determine the effects of minimum neighborhood size on the predictive maps. Three predictive maps of the model selected using default settings, with minimum neighborhood sizes of 1, 3, and 10, were created in Hyperniche and compared qualitatively.

Model Testing

The model created in Hyperniche using the suggested default settings provided a predictive map with the probability of species occurrence ranging from 0-100%. To better provide areas for field model validation, species occurrence probabilities were classified into three groups: Low (0-32%), medium (33-65%), and high (66-100%) probability. Thirty-three random points were selected in each of these probability classifications before entering the field.

Each accessible field validation point was visited. Some points were not accessible because of rugged terrain and consequently were not sampled. At each validation point, 61-m line transects were walked in four directions (north, south, east, and west) from the randomly
selected validation point. If *E. umbellatum* was seen along one of the transects, the site was designated as present, if the species was not seen along any of the transects, it was designated absent. To estimate the accuracy of the model created, the percentage of sites with *E. umbellatum* present were calculated for the low, medium, and high probability areas.

**RESULTS**

The analysis that used a minimum average neighborhood size of 40 produced a model using elevation, slope, land cover, and soil saturated hydraulic conductance (ksat) predictor variables ($\log \beta = 24.888$). Twenty-two response variable sites were in neighborhood sizes smaller than 40 and were thus excluded from the model. In the sites with *E. umbellatum* present, 87 probability estimates were improvements (higher than the naïve model) and 48 were not improvements (lower than/equal to the naïve model, or errors of omission). Where *E. umbellatum* was absent, 477 sites were improvements (lower than the naïve model) and 165 sites were not improvements (higher than/equal to the naïve model or errors of commission). The overall improvement was 72.6% (564 of 777 sites). The analysis with the neighborhood size of 100 produced a model ($\log \beta = 23.842$) with the following predictor variables: elevation, annual maximum temperature, soil electrical conductivity at a depths of 0-15, 15-25, 25-50, and 50-100 cm, soil saturated hydraulic conductance (ksat), soil organic matter at the surface and 50-100 cm, and soil pH at the surface. In this model, 20 response variable sites were excluded. Of the sites where the species were present, 96 were improvements and 38 were not improvements. Of the sites where the species were absent, 470 were improvements and 175 were not improvements. The overall percent improvement was 72.7%, with 566 of 779 sites considered improvements. Despite the slightly higher number of improvements in the model with a neighborhood size of
100, the model created with the minimum average neighborhood size of 40 had the higher log $\beta$ value, fewer predictor variables (4 compared to 10), and more complete distribution in the predictive map of the county with more high probability areas of species occurrence. Because of the higher log $\beta$ value of the model with the average neighborhood size of 40, that model was selected to create the predictive map for the field model testing phase instead of the model with the average neighborhood size of 100.

The qualitative comparison of the three predictive maps of the subset area within Tooele County with different minimum neighborhood sizes showed that the minimum neighborhood size value of 1 mapped the probability of occurrence of *E. umbellatum* for the largest area. As the minimum neighborhood size increased, the mapped area decreased (Fig. 3). However, the areas common to the maps had the same probability values, indicating that changing this parameter simply changed the extent of the predictive maps created.

The predictive map area consisted of 44.5% low probability areas, 2.5% medium probability areas, 0.1% high probability areas, and 52.9% were areas not mapped because the probability was below the minimum neighborhood size of 1. Of the 99 random locations selected for field testing, only 81 sites were accessible (27 low, 25 medium, and 28 high) and all were located on the eastern half of the county. Based on the probability ranges of the predictive map (Fig. 4), it was expected that low probability areas have species presence in 0-32% of the sites, medium to have presence in 33-65% of the sites, and high to have presence in 66-100% of the sites. In the 27 low probability field validation sites, *E. umbellatum* was present in 2 (7.4%). Of the 25 medium probability sites, the species was present at 3 (12.0%) and of the 28 high probability sites, the species was present at 8 (28.6%). Between the random test locations, *E. umbellatum* was also sampled coincidentally at 1 medium probability and 3 high probability
locations. At nearly every sample site where the species was present, it was located on a south-facing steep talus slope. The majority of locations with the species present were also in the northern half of the county. Figure 5 shows the distribution of sampling sites on the predictive model of Tooele County.

**DISCUSSION**

The model selected for field testing provided some indication of *E. umbellatum* distribution in Tooele County, although lower than anticipated *a priori*. This was particularly evident in high probability areas, which had higher *E. umbellatum* occurrence than the medium and low probability areas. A seed source producer could have used the predictive map to locate new populations of the species by beginning the investigation in the high probability areas, particularly if the producer were to begin looking in the northern half of the county. A focused effort to locate and collect seed in those high probability areas could potentially save the seed source producer time and monetary resources. In fact, the author expected the species to be present (based on personal experience) in several low and medium probability areas sampled, but it was not. Conversely, the high probability areas where the species was present were not where the author would have anticipated or focused seed collection efforts. Based on these observations, it is safe to say that the predictive model was better at locating potential seed source communities than the author’s professional experience.

It is difficult to assess the effectiveness of the model in predicting the success of an *E. umbellatum* seeding in a particular area in Tooele County based on the model. The high probability areas where the species were present were in steep talus areas that would be impossible to seed by any means other than aerially. It is not safe to assume, based on the low
frequency of species occurrence, that a particular probability would indicate a seeding success in this particular study area. However, if the model were to be applied to other localities within Utah, high probability areas may emerge with site characteristics conducive to effective rangeland seeding methods like rangeland drill seeding.

The model created during the calibration phase of analysis, because it was created from state-wide data, can be further applied to all of Utah. This was impractical for the purpose of this study due to the large field verification area and long period it took to create the Tooele County predictive map. The Hyperniche program would not allow predictive maps much larger than Tooele County, so a state-wide map would have to be compiled from predictive maps of each of Utah’s 29 counties. It would take approximately another year to produce all 29 county maps.

The evidence mentioned above that *E. umbellatum* was distributed mainly in the northern half of the county indicates that it may have been limited to the northern region by some physical barrier. Seeds of this species are small and not adapted morphologically for broad wind distribution. Neither do they have fleshy seed coats that enhance distribution by birds or mammals. It is likely that seeds are only distributed locally by gravity, water transport, erosion, or rodents. After careful examination of the field verification data, it became evident that species presence was exclusive to a single watershed, although high probability regions outside the watershed in the northern mountains (Stansbury Mountains) were not sampled in this study (Fig. 6). However, it seems that the southern Sheeprock Mountains might be forming a species barrier preventing the populations on the northeastern slopes from dispersing to the potential habitats on the southwestern slopes.

It was suspected that the artificial separation of high, middle, and low probability regions might have contributed to the lower success of the model and predictive map than anticipated.
However, upon careful assessment of field sites, the range of probabilities of the high probability present sites was the same as the high probability absent sites.

The low success of the predictive model might be related to the input data used in the study. Although the data integrity of the UDWRRT data is quite reliable and sites are generally located within homogeneous habitat patches that represent the surrounding community, occasionally sites are located in multiple patches due to landscape heterogeneity. In some cases, the UDWRRT monitoring sites span two or more community types along the 152.4-m transect and a particular species may be in a different community or niche than where the GPS location is recorded. This is generally not the case, but may have occurred enough that it influenced the predictor variables selected during the calibration phase of the model building process. It is also quite probable that $E.\ umbellatum$ was omitted from several UDWRRT locations because of sampling methodology. The UDWRRT method samples an approximately 4,645.15-m$^2$ area by actually sampling 25.0m$^2$ divided equally among 5 transects (100, 0.25-m$^2$ quadrats), approximately 0.53% of a site’s area (compared to sampling 222.97m$^2$, or 1.91% of the area, of the model field test site). There are no other means in the UDWRRT methodology to sample forb species that may fall between the transects, and would thus be omitted from the species list at a particular site.

In using the UDWRRT and herbarium specimens, it was assumed that all plant identification was correctly performed for the presence or absence of $E.\ umbellatum$ at each response variable location. However, varieties of $E.\ umbellatum$ are morphologically similar to $Eriogonum\ heracleoides$ Nutt., which belongs to a similar community complex as $E.\ umbellatum$. It is possible that these two species were confused at a UDWRRT site and $E.\ umbellatum$ falsely identified as $E.\ heracleoides$, or vice versa, which could also decrease the
integrity of the model. Welsh et al. (1993) recognizes four varieties of *E. umbellatum* in Utah, three of which had been identified in Tooele County. However, the UDWRRT data only recognizes the species taxon, so it is quite possible that the UDWRRT species represents multiple varieties from multiple habitats across Utah. This loss of taxonomic resolution could have confounded the model while identifying the predictor variables and tolerances during the calibration phase of model creation.

The field tested model selected the elevation, slope, land cover, and soil saturated hydraulic conductance (ksat) variables as those best suited to predict species occurrence. All high probability sites with *E. umbellatum* present had similar elevation, slope, and land cover (plant community) characteristics. All high probability sites with the species present were within the 1984-2280 m elevation range and were on steep (27-37°) talus south-facing slopes within mixed mountain big sagebrush (*Artemisia tridentata* ssp. *vaseyana* (Rydb.) Beetle) and curl-leaf mountain mahogany (*Cercocarpus ledifolius* Nutt.) communities. The narrow ecological variable range selected by the model strongly indicates that we were predicting species occurrence for a single variety of *E. umbellatum*. Predictions for the species taxon (multiple varieties of *E. umbellatum*) would have included a broader range of plant communities in the model thus expanding the range of species occurrence within the predictive map area.

The majority of the response variable sites were slightly biased toward mule deer winter range because the UDWRRT sites generally monitored this habitat type. Of the 799 response variable sites, 556 (69.6%) of them are located within mule deer winter range. This bias could likely impact the effectiveness of the model because only 3.8% of Tooele County is classified by the Utah Division of Wildlife Resources as mule deer (*Odocoileus hemionus*) winter range. This
could be prove to be different in counties where winter range provides a larger percentage of the county area.

Differences in scale between the input response variable data (broad-scale) and that of the output prediction data (finer scale) may have contributed to decreased accuracy for the predictive map. Only 65 of the 799 (8.1%) input response variable sites were in Tooele County. Although the model did provide several high probability areas with species occurrence, it also provided several high probability areas where no *E. umbellatum* occurred (e.g. the area south of the Sheeprock Mountains). The lack of representative input data for the entire county may have changed the model calibration process and contributed fewer high probability areas with no *E. umbellatum* present.

The qualitative comparison of the subset maps with different neighborhood size values (used when creating the predictive map) indicated that increasing the neighborhood size restriction when creating the map only decreased the map area. This restriction would only be desirable when the user desires to ignore areas under a certain probability criterion. In the case of this study, using 1 (the lowest neighborhood size), was the desirable setting in that it mapped all probabilities greater than 0. The minimum neighborhood size criterion (positive integer) limits small probability areas that a researcher may desire to display on a map for understanding where there is no chance of species occurrence, but may not be altogether useful ecologically. With the current setting, there is nothing to distinguish probabilities that fall between 0 and 1.

Based on the two models compared in the study, it seems that with multiplicative regression analysis, models with more predictor variables (and predictor variables with missing data values) create maps with larger areas of missing data (probabilities of 0). This is due to the modeling approach, where a higher number of predictors increase the likelihood that a predictor
with a 0 probability value will be included in the predictive mapping of a particular area. When
the variable with a 0 probability is multiplied by the predictor variables, the mapped predictive
value is 0, and is left off the map. Thus we suggest, based on results of this project, that the
number of variables selected in a model be considered when selecting a model along with the log
\( \beta \) value. After further research comparing the trade-offs of predictor variables and log \( \beta \) values,
perhaps an index could be developed to further aid researchers in selecting the best models.

In the end, the predictive map created using NPMR did provide locations where *E. umbellatum*
occurs naturally within Tooele County, although the percentage of sites with the
species present rate of success was lower than anticipated for the high and medium probability
areas. This method of locating seed source populations remains much less time consuming than
an undirected approach. The species may have been absent in many areas of Tooele because of a
limited species distribution rather than an unfavorable ecological niche. Further research
involving *E. umbellatum* seeding success in high probability areas in Utah might provide more
information on this question.

**LITERATURE CITED**

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological

research Journal of Forestry 64:691.


Monsen, S.B. 2004. Restoration or rehabilitation through management or artificial


Fig. 1. Relief map and location of Utah in the United States. The area within the white border is Tooele County.
Fig. 2. Map of the SSURGO soil data coverage for the state of Utah. The solid gray areas are those with soil data coverage.
Fig. 3. Comparison of the subset predictive maps with map creation input minimum neighborhood sizes of 1, 3, and 10. White areas are those with no predicted species occurrence, light gray are those with low probabilities, those with dark gray are medium probabilities, and black are those with high probabilities. The input minimum neighborhood size of the map on the left is 1, the center is 3, and right is 10. As the minimum neighborhood size increased, the area displayed on the predictive map decreased.
Fig. 4. Predictive map of *Eriogonum umbellatum* within the study area. The predictor variables selected during the analysis of this model included elevation, slope, land cover, and soil saturated hydraulic conductance (ksat). The areas with no predicted species occurrence are displayed in white, those with low probability in light gray, those with medium probability in dark gray, and those with high probability in black.
Fig. 5. Distribution of field verification sample sites across the Tooele County predictive map.
Fig. 6. Relief map of the study area with field verification sample sites and watershed boundary. The red points are those with expected low probability of occurrence, those with yellow are medium, and those with green are high. Circular points are sites with species absence and star points are species presence. The blue line represents the