Iterative Distribution Modeling for Two Endemic Plants of the Northern Piceance Basin

prepared for:
U.S. Fish and Wildlife Service

Karin Decker, Michelle Fink, Jill Handwerk, and Gabrielle Smith

Colorado Natural Heritage Program
Colorado State University
Campus Delivery 1475
Ft. Collins, Colorado 80523-1475

April 2013

Table of Contents

Introduction 2
Species distribution modeling 2
Study elements 3
Methods 4
Input data 4
Maximum entropy models 6
Field testing 7
Model revision 8
Discussion 14
References 16
Appendix A: List of 1:24,000 geologic quads digitized 18
Appendix B: Side-by-side comparison of models 20

Figure 1: Study area and documented occurrences 5
Figure 2. Predicted suitable habitat for *Physaria congesta*, Maxent model. 10
Figure 3. Predicted suitable habitat for *Physaria obcordata*, Maxent model. 10
Figure 4. Predicted suitable habitat for *Physaria conjesta*, random forest models. 13
Figure 5. Predicted suitable habitat for *Physaria obcordata*, random forest models. 14
Figure 6. Comparison of predicted high suitable habitat acreage for all models. 15

Table 1: Environmental variables used for modeling. 5
Table 2. Variables selected in 2011 draft models 7
Table 3. AUC and AICc for Maxent models. Selected models are shaded. 9
Table 4. Predictive variables included in selected Maxent models. 11
Table 5. Relative importance of variables in random forest replicate models. 12
Table 6. Comparison of predicted suitable habitat acreage for all models. 14

cover photo by Jill Handwerk

INTRODUCTION

Species distribution modeling

Species distribution modeling is one of many tools available to assist managers in understanding the potential distribution of rare and endemic species when regulating and prioritizing different land-use scenarios. Developing a predictive model of the distribution of a particular species can involve several different techniques, and be reported under a variety of names. All such models, however, are based on the ecological principle that the presence of a species on the landscape is controlled by a variety of biotic and abiotic factors, in the context of biogeographic and evolutionary history. Because we rarely, if ever, have complete and accurate knowledge of these factors and history, we can only seek to predict or discover suitable habitat by using characteristics of known occurrences of the taxon in question.

The modeling process is further constrained by our inability to measure habitat characteristics accurately on a continuous spatial scale. As a result, modeling factors are usually an approximation of the environmental factors that control species distribution, using available data that is probably only a surrogate for the actual controlling factors. In the context of our study, species distribution modeling is a process that uses a sample of a real distribution (known locations or element occurrences) to build a model (estimate) of suitable environmental conditions (and, by implication, unsuitable conditions), and map that model across a study area. In this study we used an iterative modeling approach to investigate the potential distribution of two rare species: *Physaria (Lesquerella) congesta,* and *Physaria obcordata.* All models used spatially referenced datasets of environmental variables (i.e., elevation, soil, geology, temperature, precipitation, and other factors). Modeling techniques are further discussed under “Methods” below.

Models for these two species were originally produced using the DOMAIN and biophysical “envelope” methods (Decker et al. 2006). This effort covered an area of about 759,000 acres (1,185 sq. miles) in the northern Piceance Basin (Figure 1). Because these models were based on very generalized geological mapping, they were regarded as fairly imprecise predictors of suitable habitat, in that a substantial amount of unsuitable habitat was included. Ideally, species distribution models are parameterized with environmental data that are known to be highly predictive of conditions determining the ability of a species to persist. Unfortunately, for many species we lack even basic life-history information that could guide input selection. Even when extremely detailed information about important microhabitat factors is available, these factors are generally not mapped or otherwise spatially represented at a scale that is equivalent to that experienced by the organism. Since the two species in this study appear to be closely tied to particular geological substrates, we decided to refine the distribution models with more precise geologic mapping.

Study elements

Physaria (Lesquerella) congesta

Physaria congesta Rollins, the Dudley Bluffs bladderpod, is an herbaceous perennial member of the Brassicaceae (mustard family) that is endemic to northwestern Colorado. The species was described by Reed Rollins (1984) from specimens collected at North Dudley Gulch, Rio Blanco County, Colorado. The entire range as currently known lies within 10 miles of this type locality. Plants are small, cushion-shaped, up to 3 cm in diameter, with a congested mass of bright yellow flowers and silvery leaves rising from a long, thin taproot. Flowering is typically during April and May, and fruit set from late May into June.

Physaria congesta is believed to be more-or-less confined to outcrops of the Thirteenmile Creek Tongue of the Green River Formation. This formation is described by Hail and Smith (1994):

> [The] Thirteenmile Creek Tongue, which is one of the most widespread tongues of Green River Formation in northern part of Piceance Creek basin, consists of mostly light-gray to white, variably silty marlstone, smaller amounts of locally oolitic, ostracodal, and algal limestone, and some sandstone, siltstone, and claystone; it includes three or four thin oil-shale beds in the southern part of the outcrop area. Thickness of Thirteenmile Creek Tongue is about 20-200 ft (6.1-61m).

Hail and Smith (1994) included the Thirteenmile Creek Tongue as the lowermost part of their group C of tongues of the Uinta and Green River Formations. Reported habitat is described as barren white shale outcrops exposed along downcutting drainages. Known occurrences are concentrated in the Piceance Creek drainage near its intersection with Ryan Gulch, and in the Yellow Creek drainage to northwest, including the Duck Creek drainage.

The Dudley Bluffs bladderpod was first listed on February 06, 1990, and is currently designated as Threatened under the Endangered Species Act. *Physaria congesta* is currently known from seven occurrences (consisting of 119 distinct mapped polygons) in the northern Piceance Basin in Rio Blanco County. The species is ranked G1S1, critically imperiled both globally and within the state, by the Colorado Natural Heritage Program (2013) and NatureServe (2012).

Physaria obcordata

Physaria obcordata Rollins, the Piceance Basin twinpod, is also a perennial member of the Brassicaceae. *Physaria obcordata* was described by Rollins (1983) from specimens collected by Baker and Naumann during a 1982 survey of the Piceance Basin. The specific epithet *obcordata* refers to the heart-shaped silique or fruit. Plants have flowering stems 12-18 cm tall that arise from a basal tuft of silvery leaves. Flowers are yellow, and typically present in May and June.

The range as currently known is slightly broader than that of *Physaria congesta*, with some occurrences being separated by as much as 26 miles. It is sympatric with *P. congesta* in the Piceance Creek/Ryan Gulch area, and also occurs in the Yellow Creek drainage. Additional occurrences are known further to the northwest in the East Fork of Spring Creek, Fletcher Gulch and Yanks Gulch drainages. Recent discoveries have also been made in a few locations on bluffs south of the White River. O'Kane (1988) reported that *P. obcordata* occurs primarily on the Thirteenmile Creek Tongue and on the Parachute Creek Member of the Green River Formation.

The main body of the Parachute Creek Member in the Northern Piceance Basin is described by Hail and Smith (1994) as:

> Dolomitic marlstone, oil shale, and some marly siltstone and sandstone, numerous thin beds of analcimized tuff, and sparse algal limestone beds. Marlstone is massive to platy, gray to light grayish brown, weathers light gray. Oil shale is thin, even bedded, locally papery, medium to dark brown, weathers light gray; richest oil shale weathers light bluish gray. Clay shale is thin bedded to papery, brown to brownish gray. Parachute Creek Member contains several zones of rich, potentially valuable oil shale, constituting by far most of the oil-shale resources in the Green River Formation. … Thickness of entire member on outcrop south of White River ranges from about 400 ft (120m) in eastern part of basin to about 1600 ft (488 m) in western part of basin. Thickness in subsurface may exceed 2000 ft (610m).

The Parachute Creek Member includes the Mahogany ledge, one of the richest oil-shale zones in the basin, as well as potentially valuable deposits of nahcolite and dawsonite. This member lies below the intertongued Uinta and Green River Formations and above the Wasatch Formation. The adjacent Garden Gulch Member beneath the Parachute Creek Member may also support some occurrences.

Physaria obcordata was listed as Threatened under the Endangered Species Act at the same time as *Physaria congesta,* and is currently known from eleven occurrences (consisting of 84 distinct mapped polygons) in the northern Piceance Basin in Rio Blanco County. The species is ranked G1G2/S1S2 (imperiled or critically imperiled both globally and in the state, at high risk of extinction due to very restricted range with 20 or fewer populations or other factors) by the Colorado Natural Heritage Program (2013) and NatureServe (2012).

METHODS

Input data

The study area for the 2011 model revision was an approximate 550,000 acre (860 sq. mile) area in the northern portion of the Piceance basin (Figure 1). This study area boundary was determined by the aggregation of 15 7.5' geologic quads that were digitized for this effort, and was about 209,000 acres smaller than the area modeled in Decker et al. (2006). This area encompassed all known occurrences of the two species in 2011, but did not necessarily include all suitable geologic substrates, particularly for *Physaria obcordata*. Models were constructed with data from known locations of the target species using element occurrence records from the Colorado Natural Heritage Program databases (Figure 1). Element occurrence records were updated and reviewed for completeness prior to modeling, to ensure that the most accurate information was used. Element presence data points were produced from mapped polygons updated in 2011. Polygon centroids were used, except in the case of a few very large polygons. For these occurrences, a 500 m grid was used to cut the polygon, and the centroid of each section selected as a sample point. This procedure was intended to insure that a more representative sample of presence values of each covariate dataset would be included. For the model revisions in 2013, we updated and revised element occurrence locations, and used the same general procedures to produce model points. We also digitized eight additional 7.5' geologic quads to expand the study area to include suitable geologic substrates omitted from the 2011 study area. The final revised study area (Figure 1) was a little over 867,700 acres (1,355 sq. miles).

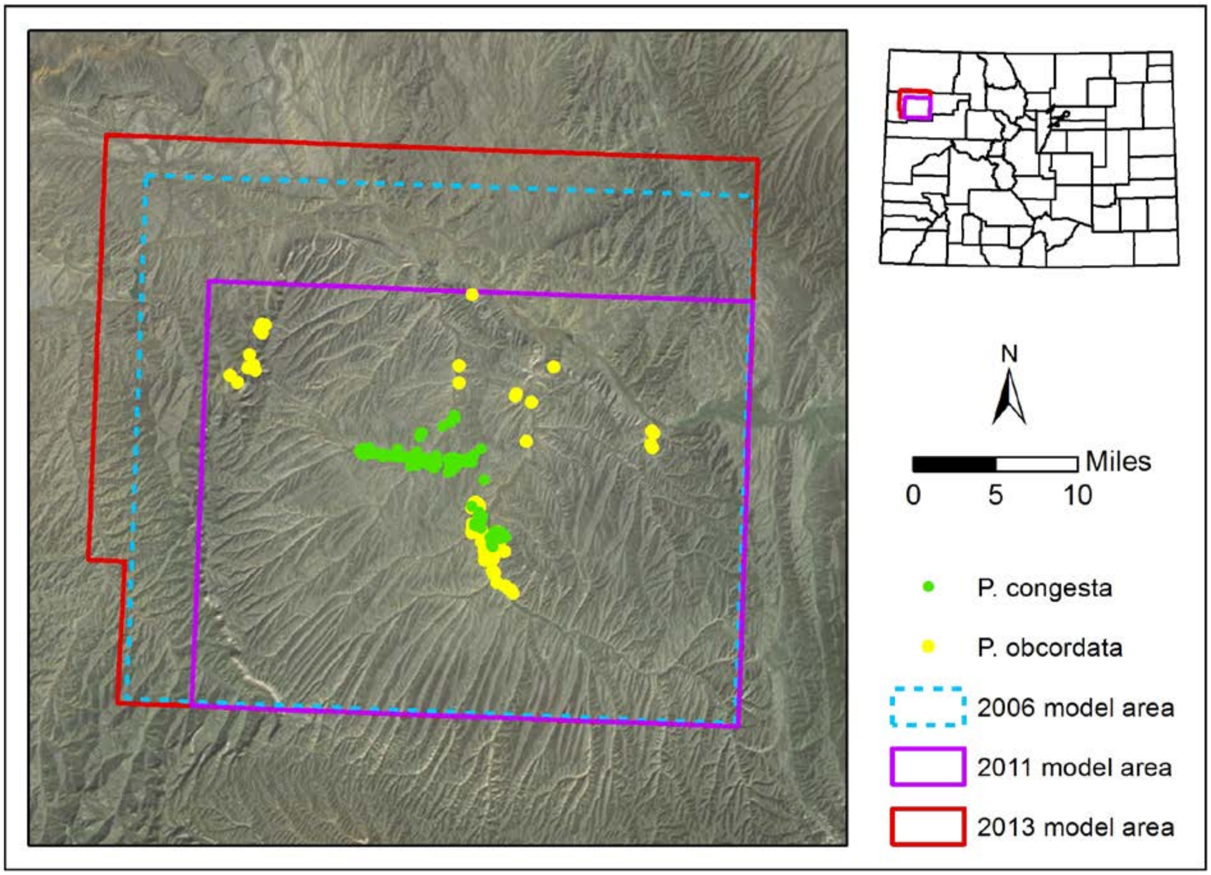


Figure 1: Study area and documented occurrences.

Environmental covariates for occurrence points were derived from digital raster data in ArcGIS 9.3-10 (ESRI 2009-2010). Datasets were processed to a common projection, clipped to the appropriate study area, and resampled as necessary to a uniform cell size. Environmental data used and sources are listed in Table 1.

Table 1: Environmental variables used for modeling.

Continuous Variables	Units	Abbreviation	Source
Elevation	m	Elevation or ned10	USGS 30m and 10m Digital Elevation Model (DEM) for Colorado
Local Relief	m	relief	Derived from DEM
Aspect (northness)	m	northness	Derived from DEM
Aspect (eastness)	m	eastness	Derived from DEM
Annual mean precipitation	cm	totppt	Daymet - Climatological summaries for the conterminous United States 1980-1997 (1km)
Spring precip (Mar-Apr-May)	cm	springppt	Daymet
Summer precip (Jun-Jul-Aug)	cm	summerppt	Daymet
Fall precip (Sep-Oct-Nov)	cm	fallppt	Daymet

Winter precip (Dec-Jan-Feb)	cm	winterppt	Daymet
Minimum air temperature for each of 12 months	°C	mintemp00	Daymet
Growing degree days – annual (average air temp above 0 °C)	degree -days	growday	Daymet
Number of frost days	days	frostday	Daymet
Distance to Green River Fm, Thirteenmile Creek Tongue, upper part (Tgtu)	m	dist_tgtu	Generated from surface geology (see below)
Distance to Green River Fm (Tg)	m	dist_tg	Generated from surface geology (see below)
Categorical Variables	**Values**		**Source**
Surface Geology		geology	Digitized onscreen from 1:24,000 scale geologic maps – see list in appendix
Soil type		ssurgo	Soil Survey Staff, NRCS, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for Rio Blanco County, Colo. http://soildatamart.nrcs.usda.gov
Vegetation type		veg	USGS National Gap Analysis Program. 2004. Provisional Digital Land Cover Map for the Southwestern United States. Version 1.0. RS/GIS Laboratory, College of Natural Resources, Utah State University.

Maximum entropy models

The maximum entropy (Maxent) modeling procedure (Phillips et al. 2004, 2006) has been widely used in species distribution modeling and performs well in comparison with other methods (Elith et al. 2011). This procedure uses the environmental covariate values from occurrence records plus 10,000 randomly selected background points to estimate a probability distribution that is consistent with data from known locations. This estimate is as close as possible (has maximum entropy) to the estimate from the background data (the null model), since, without any data, we would have no reason to think that the species would be more likely to be in one location than any other. Species distribution is estimated by minimizing the distance between the occupied and background, subject to constraining the means of estimated occupied factors to be close to observed mean. Constraints ensure that the mean for a variable in the estimated distribution is close to the mean across the locations with occurrences. The raw solution is transformed to logistic output, becoming more-or-less a probability estimate of occurrence.

For the 2011 revision of the *Physaria* models we used Maxent software (version 3.3.3e-k, Dudik et al. 2010) to generate probability surfaces for the two species. Twenty-five percent of the presence points were withheld by the procedure for testing. The 2011 draft models were produced at a resolution of 30m x 30m, based on the 30m DEM. Several iterations were run for each species with some highly correlated environmental covariates removed. AUC values ("area under curve," commonly used to evaluate model performance) of model sets for both species were the same to two significant digits for both training and test data. Because all model iterations appeared equally fit, models for field testing were selected by comparing the predicted suitable habitat with mapped polygons of known occurrences.

Field testing

For both species the 2011 model with the fewest environmental parameters also had the best coverage of occurrence polygons and was used for field testing. Suitable habitat for *Physaria congesta* is strongly tied to the Thirteenmile Tongue geology, while habitat for *P. obcordata* also incorporates climate factors (Table 2).

Table 2. Variables selected in 2011 draft models.

Species	Variable	Percent contribution
Physaria congesta	Distance to13mile Tongue	64.6
145 presence pts used	Elevation	14.8
	April min. temp	10
	Summer precip.	6.6
	Surface geology	2.8
	Growing Degree Days	1.2
Physaria obcordata	Growing Degree Days	26.2
93 presence points used	Distance to Green Riv. Fm	23.4
	Surface geology	22.1
	Annual precip.	13.1
	April min. temp	7.8
	Vegetation type	7.3

Draft models were produced in the spring of 2011. The RRQRR (Reversed Randomized Quadrant-Recursive Raster, Theobald et al. 2007) technique was used to create spatially balanced random sampling points for the modeling area. GIS datasets of random points for each species were provided to field surveyors. Surveyors did not have the model available during field visits. Surveys to ground-truth the random points took place in mid-May and late June of 2011. The goal was to visit a minimum of 30 points in suitable habitat for each species, but searchers always looked for both species at a point. Only points within 750 meters of existing roads were chosen for sampling to facilitate access to the sampling sites. A total of 67 random points in areas of high probability for *Physaria congesta* were visited with 2 positive results. Nineteen (28%) of the random points were identified as potentially suitable habitat by field observations, three additional sites were near suitable habitat, and the remaining points were not in suitable habitat for the species. Thirty-five random points in modeled areas of high probability for *Physaria obcordata* were visited. No *P. obcordata* were found, but *P. congesta* was found at one point. Of the 35 points visited, only 6 (17%) fell in areas identified in the field as potentially suitable habitat; the remaining points were in unsuitable habitat for the species.

Maxent model revision

We thought that the poor field validation results for the 2011 models could be due in part to the size of the data cells used, which is still quite large in comparison to the size of the plants. In addition, two new *Physaria obcordata* occurrences were found near the northern extent of the current habitat models, indicating that the modeling of an enlarged study area would be beneficial. We decided to try to improve model resolution for the next iteration by acquiring larger scale data for two of the primary input sources: geology and the Digital Elevation Model (DEM). To date, attempts by the BLM in conjunction with CNHP to acquire higher resolution geology maps of the study area from industry contacts have been unsuccessful. In the interim, CNHP digitized additional 1:24000 scale geology quads to include in an expanded habitat model and acquired the 10 meter DEM.

Five Maxent models were run for each species in 2013:

1. The first model included all 28 variables (Table 1).
2. All variables that had a percent contribution >=1 or permutation importance of >=0.5 in the first model.
3. Variables from the first model that appeared to be the most influential in jackknife permutations, plus the most influential of the monthly minimum temperature variables.
4. All variables from the initial model that had any non-zero percent contribution, except those identified by jackknife as consistently the least important.
5. All variables from previous runs that appeared to be consistently the most important, either by percent contribution, or by jackknife permutation.

Results from the five model runs were compared by calculating the Akaike information criterion, corrected for finite sample size (AICc), using the program ENMTools, version 1.3 (Warren et al. 2010, Warren and Seifert 2011). For each species the model with the lowest AICc value was selected.

Random forest models

During the 2013 model revision process, we also used the random forests technique to generate predictive models, as a comparison to Maxent. Random forests (Breiman 2001) is a classification and regression tree (‘‘CART’’) technique. CART analyses use a variety of algorithms for predicting continuous or categorical variables from a set of continuous or categorical effect variables. Regression-type analyses generally attempt to predict the values of a continuous variable and classification-type analyses attempt to predict values of a categorical dependent variable (class, group membership, etc.). A simple binary classification-type analysis predicts the presence or absence of a species according to the values of various environmental factors. At each iteration, the recursive CART process determines which environmental variable and value best divides the set of all points into a “mostly present” and “mostly absent” set. The final result is a dichotomous tree showing the conditions of each split that describe suitable (present) and unsuitable (absent) environments. The random forest technique grows many such trees from a data set, using a randomized subset of predictors, and combines the predictions from all the trees (Prasad et al. 2006, Cutler et al. 2007). Because a large number of trees (a “forest”) are grown, the technique is robust against over fitting (Breiman 2001).

Random forest models were tried first with true absence records from ground-truthing and other field survey work in the area. Absence points identified during field work as apparently suitable were omitted. Models were also run with randomly generated psuedo-absenses falling outside mapped occurrences. Following the recommendations of Barbet-Massin et al. (2012), we used the same number of psuedo-absences as there were presence points, and averaged 10 runs. Random forest was implemented in R (R Core Team 2012), using packages randomForest (Law and Wiener 2002), raster (Hijmans and van Etten 2012), rgdal (Bivand et al. 2013), and sp (Pebsma and Bivand 2005, Bivand et al. 2008), using code written by Jeffrey Evans (Evans 2012) to select model parameters at a parsimony level of 0.3. Resulting models were predicted in ascii format, converted to ESRI raster, and averaged.

RESULTS

Maximum entropy

All models had very high AUC (area under curve) values for both training and test data. The AICc (Akaike information criterion, corrected) values for the different models of *P. congesta* were fairly closely grouped; those for *P. obcordata* were more variable (Table 3). The selected models with the lowest value of AICc (Figures 2 and 3) indicate the probability of suitable habitat for each raster cell.

Table 3. AUC (area under curve) and AICc (Akaike information criterion, corrected) for Maxent models. Selected models are shaded.

Species-model	AUC Train	AUC Test	AICc
Physaria congesta 1	0.993	0.988	4117.995
Physaria congesta 2	0.991	0.986	3735.152
Physaria congesta 3	0.991	0.987	3752.047
Physaria congesta 4	0.992	0.988	3788.760
Physaria congesta 5	0.991	0.989	3802.458
Physaria obcordata 1	0.993	0.988	17785.790
Physaria obcordata 2	0.993	0.988	3437.032
Physaria obcordata 3	0.992	0.990	n/a
Physaria obcordata 4	0.993	0.989	3050.636
Physaria obcordata 5	0.989	0.982	2792.797

Variables used in the selected models appear in Table 4. Suitable habitat for *Physaria congesta*, as expected, is closely tied to the Thirteenmile Tongue of the Green River Formation, which accounts for nearly 70% of model variability. This species is also apparently more reliant on tolerable temperatures than precipitation. Elevation is a contributing factor as well, presumably due to the narrow band of exposure of the favored substrate.

Suitable habitat for *Physaria obcordata* is also tied to geologic substrate, but with less specificity than for *P. congesta*. Winter minimum temperatures appear to constrain habitat, as well as seasonal precipitation patterns (Table 4).

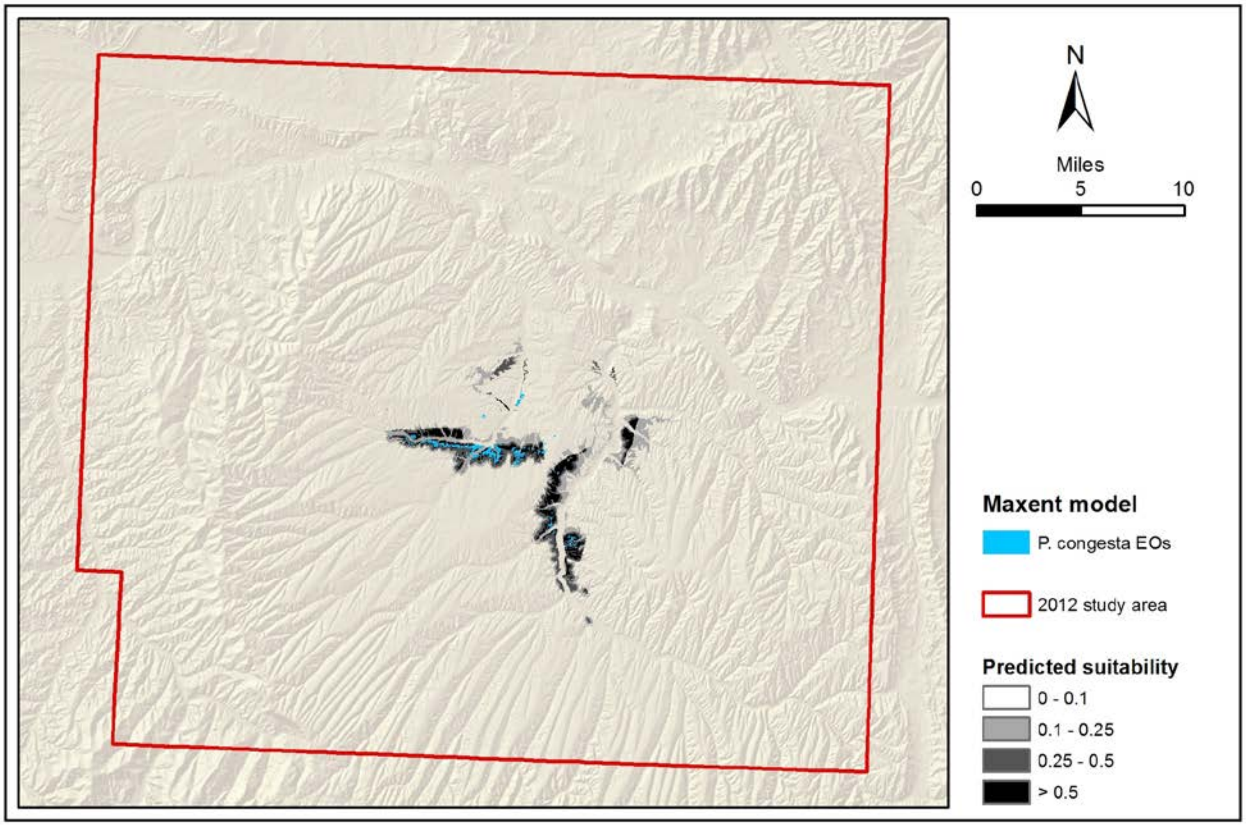


Figure 2. Predicted suitable habitat for *Physaria congesta*, Maxent model with lowest AICc.

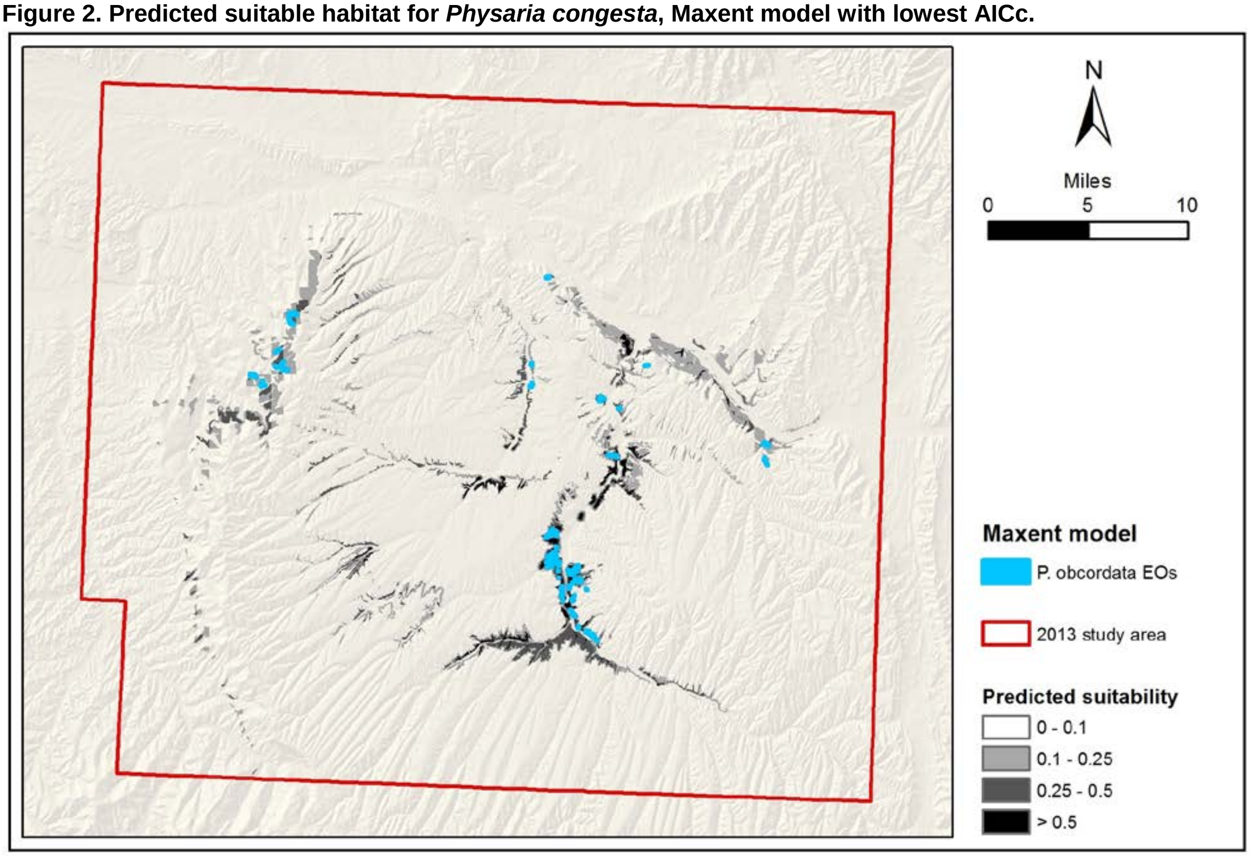


Figure 3. Predicted suitable habitat for *Physaria obcordata*, Maxent model with lowest AICc.

Table 4. Predictive variables included in selected Maxent models.

Species	Variable	Percent contribution	Permutation importance
Physaria congesta	Distance to13mile Tongue	69.5	94.1
143 presence pts used	Jan min temp	8.8	0.8
	Elevation	7.6	1.8
	Oct min temp	6.4	2.1
	Surface geology	4.0	1
	March min temp	3.6	0
	June min temp	0.1	0
	Growing Degree Days	0	0.2
Physaria obcordata	Distance to Green Riv. Fm	33.8	57.4
89 presence pts used	Soil type	28.9	3.6
	Feb min temp	22.3	21.6
	Summer precip	7.5	2.1
	Spring precip	5.1	2.4
	Jan min temp	1.7	8.1
	Fall precip	0.6	0
	Winter precip	0.2	4.8

Random forest

Random forest models that used pseudo-absences had an average misclassification rate of 1.88% for *Physaria congesta*, and 7.64% for *P. obcordata*. Misclassification rates for models using true absence points were higher, at 9.01% for *P. congesta*, and 11.27% for *P. obcordata*. Examination of the geospatial output suggests that the pseudo-absence models are somewhat under fit, including habitat where the species is known to be absent, while the true-absence models may be over fit, leaving out some habitat where the species is known to occur (Figures 4 and 5).

Distance to the Green River Formation, Thirteenmile Creek Tongue, upper part (Tgtu) and elevation were consistently the most important variables for *P. congesta*, with distance to Tgtu always most important, and elevation second. Other variables chosen in all models were growing degree days and distance to Green River Fm (Tg); January minimum temperature was omitted from a single model. Eight of ten runs with pseudo-absences chose only seven variables, and the choice included spring precipitation, total annual precipitation, or May minimum temperature. True-absence *P. congesta* runs chose either 7, 14, or 20 variables (Table 5).

Physaria obcordata models, while also closely tied to distance from both Thirteenmile Creek Tongue, upper part (Tgtu) and Green River Fm (Tg), appeared to be influenced by winter minimum temperatures, and either relief or generalized geological substrate. The importance of relief or geologic substrate presumably reflects the tendency for the species to occur on steep, erodible slopes of Green River Formation members.

Table 5. Relative importance of variables in random forest replicate models.

Physaria congesta

Pseudo-absence models			True-absence models		
Variable	Mean importance	# times selected	Variable	Mean importance	# times selected
dist_tgtu	1.0000	10	dist_tgtu	1.0000	10
ned10m	0.2819	10	ned10m	0.4976	10
growday	0.1560	10	mintmp01	0.4805	10
dist_tg	0.1495	10	dist_tg	0.4135	10
mintmp01	0.1416	9	mintmp02	0.3779	8
geology	0.1388	10	mintmp10	0.3752	9
springppt	0.1167	6	mintmp03	0.3652	10
mintmp05	0.1162	5	growday	0.3595	9
totppt	0.1033	4	summerppt	0.3421	10
fallppt	0.0968	2	mintmp12	0.3323	8
mintmp04	0.0944	2	mintmp09	0.2987	8
mintmp03	0.0867	2	mintmp04	0.2682	7
mintmp02	0.0773	2	mintmp06	0.2483	8
mintmp12	0.0761	1	totppt	0.2462	6
mintmp11	0.0667	1	fallppt	0.2322	7
eastness	---	0	mintmp05	0.2022	4
frostday	---	0	springppt	0.1896	4
mintmp06	---	0	mintmp07	0.1885	3
mintmp07	---	0	geology	0.1871	3
mintmp08	---	0	mintmp08	0.1825	4
mintmp09	---	0	mintmp11	0.1724	2
mintmp10	---	0	eastness	---	0
northness	---	0	frostday	---	0
relief	---	0	northness	---	0
ssurgo	---	0	relief	---	0
summerppt	---	0	ssurgo	---	0
veg	---	0	veg	---	0

Physaria obcordata

Pseudo-absence models			True-absence models		
dist_tgtu	1.0000	10	ned10m	1.0000	10
dist_tg	0.5305	10	relief	0.8298	10
geology	0.3332	10	dist_tg	0.6073	10
ned10m	0.2735	10	dist_tgtu	0.5824	10
mintmp02	0.2575	10	mintmp01	0.5530	10
mintmp01	0.2334	10	mintmp02	0.4366	10
mintmp12	0.2216	1	mintmp12	0.3389	10
springppt	0.2127	8	mintmp03	0.2239	1
totppt	0.2014	1	fallppt	0.2168	1
eastness	---	0	geology	0.2136	1
fallppt	---	0	mintmp11	0.1949	1
frostday	---	0	growday	0.1664	1
growday	---	0	mintmp10	0.1644	1

Pseudo-absence models			True-absence models		
Variable	Mean importance	# times selected	Variable	Mean importance	# times selected
mintmp03	---	0	mintmp04	0.1631	1
mintmp04	---	0	eastness	---	0
mintmp05	---	0	frostday	---	0
mintmp06	---	0	mintmp05	---	0
mintmp07	---	0	mintmp06	---	0
mintmp08	---	0	mintmp07	---	0
mintmp09	---	0	mintmp08	---	0
mintmp10	---	0	mintmp09	---	0
mintmp11	---	0	northness	---	0
northness	---	0	springppt	---	0
relief	---	0	ssurgo	---	0
ssurgo	---	0	summerppt	---	0
summerppt	---	0	totppt	---	0
veg	---	0	veg	---	0

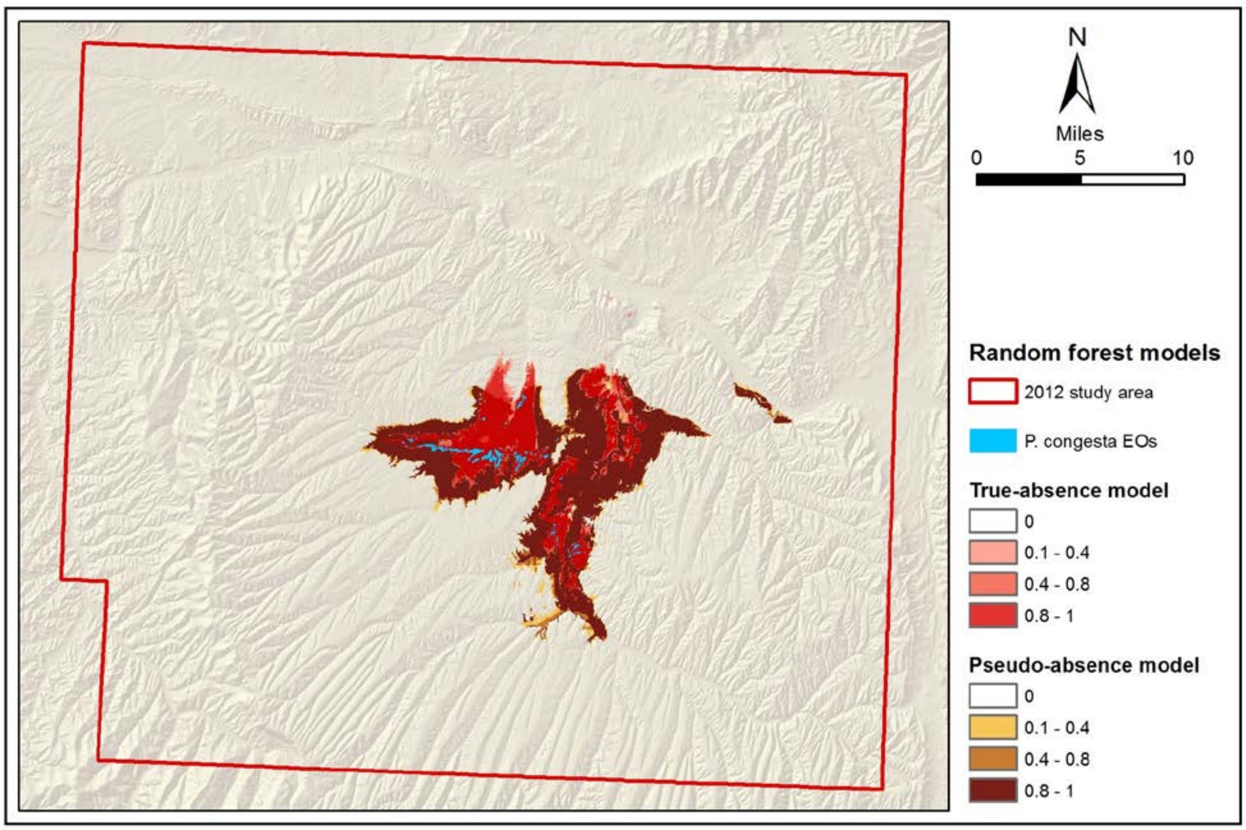


Figure 4. Predicted suitable habitat for *Physaria congesta*, random forest models.

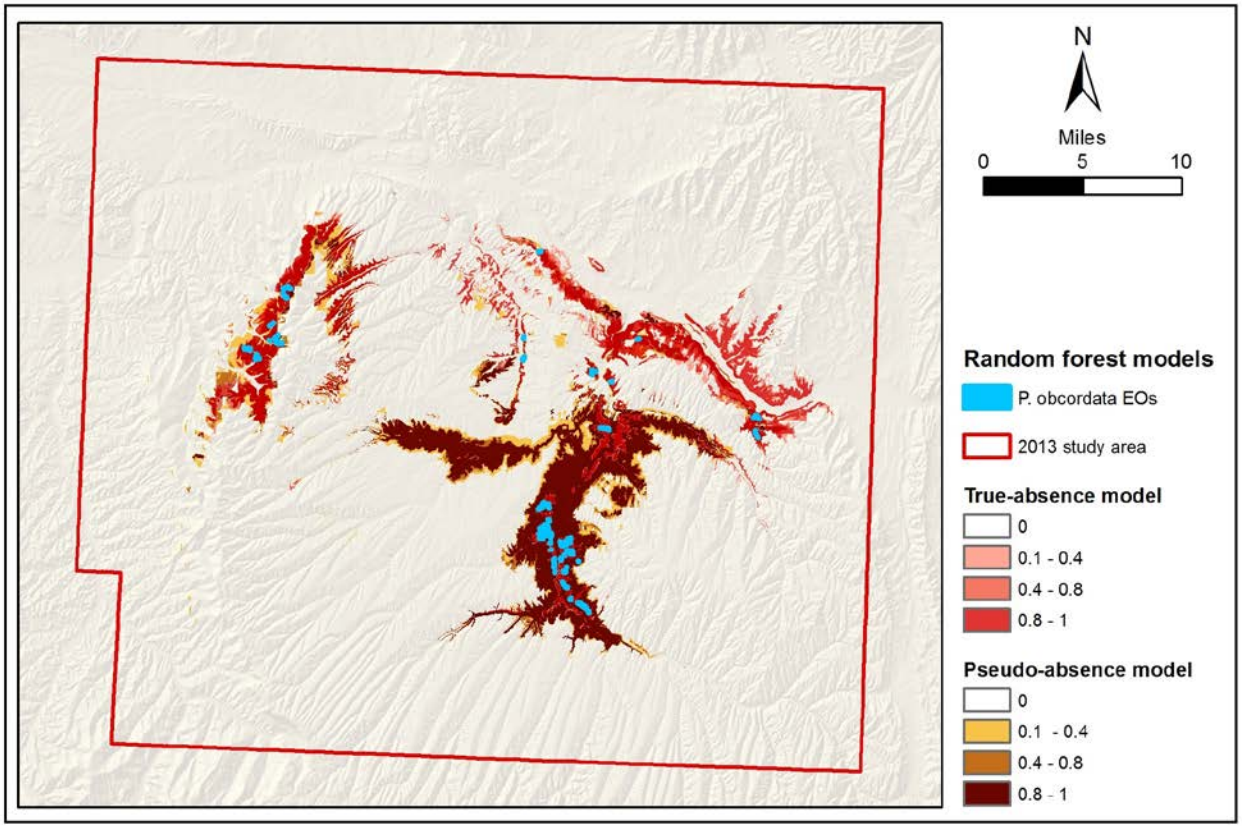


Figure 5. Predicted suitable habitat for *Physaria obcordata*, random forest models.

DISCUSSION

Although we were able to narrow our predictions for potential suitable habitat for both species through the iterative process (Table 6, Figure 6), we appear to have reached a limit for model precision with currently available data. No models were 100% successful in predicting both known presence and known absence. Furthermore, because points are used in the modeling procedure, even “core” areas (5,240 ac for *P. congesta* and 5,600 ac for *P. obcordata*) that were predicted by all three models did not include the entire extent of all mapped occurrences.

Table 6. Comparison of predicted suitable habitat acreage for all models. RF PA = random forest pseudo-absence, RF TA = random forest true-absence.

		Physaria congesta		*Physaria obcordata*	
		Acres of suitable habitat		Acres of suitable habitat	
Year	Model type	High probability	Moderate probability	High probability	Moderate probability
2006	Domain	73,765	15,364	149,326	22,022
	Envelope	73,916	na	141,438	na
2011	Maxent draft	17,692	19,617	33,409	58,478
2013	Maxent	8,231	7,149	12,087	44,281
	RF PA	46,012	2,648	58,628	15,742
	RF TA	15,721	2,316	31,971	6,896

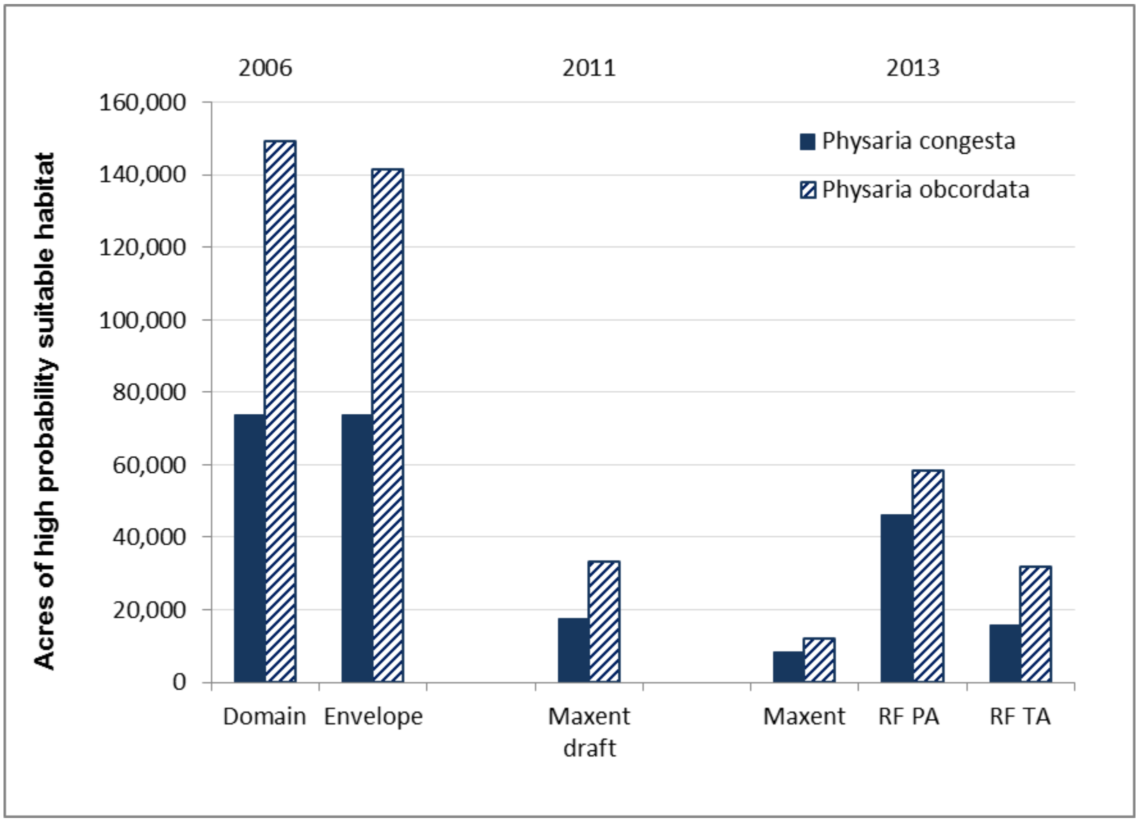


Figure 6. Comparison of predicted high suitable habitat acreage for all models.

Acreage of a "core" area predicted by all three models was notably smaller than that predicted by any of the individual models. It appears likely that there is anywhere from 8,000 to 15,000 acres of potential habitat for *Physaria congesta* and perhaps 12,000 to 32,000 acres for *P. obcordata*. Notes from 2011 field work indicate that there are patches of the appropriate geologic substrate(s) that are outside documented occurrences, and which are too small to have been mapped at the 1:24,000 scale.

How shall we interpret these uncertainties? As always, models represent a best-guess scenario in which we know that our data is incomplete and of insufficiently fine resolution. Although these models can not completely replace field verification, they can focus inventory effort on the most likely areas to support these species.

REFERENCES

Barbet-Massin, M., F. Jiguet, C.H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3:327-338.

Bivand, R.S., E.J. Pebesma, and V. Gomez-Rubio. 2008. Applied spatial data analysis with R. Springer, NY. http://www.asdar-book.org/

Bivand, R., T. Keitt, and B. Rowlingson. (2013). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.8-5. http://CRAN.R-project.org/package=rgdal

Breiman, L. 2001. Random Forests. Machine Learning 45:5-32.

Colorado Natural Heritage Program [CNHP]. 2011-2013. Biodiversity Tracking and Conservation System (BIOTICS). Colorado Natural Heritage Program, Colorado State University, Fort Collins, CO.

Cutler, D.R., T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

Decker, K. A. Lavender, J. Handwerk, and D.G. Anderson. 2006. Modeling the potential distribution of three endemic plants of the northern Piceance and Uinta Basins. Colorado Natural Heritage Program, Colorado State University, Fort Collins, CO. Available: http://www.cnhp.colostate.edu/download/documents/2006/PEGR_LECO_PHOB_report.pdf

Dudik, M., S.Phillips, and R. Shapire. 2010. Maximum Entropy Modeling software, versions 3.3.3e and 3.3.3k. Available: http://www.cs.princeton.edu/~schapire/maxent.

Elith, J., S.J. Phillips, T. Hastie, M. Dudik, Y.E. Chee, and C.J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43-57.

ESRI. 2009-2010. ArcGIS 9.3-10 SP3. ESRI, Redlands, CA.

Evans, J. 2012. rf.modelSel.R. Available: http://conserveonline.org/workspaces/emt/documents/random-forests-model-selection/view.html

Hail Jr., W.J., and M.C. Smith. 1994. Geologic map of the northern part of the Piceance Creek Basin, northwestern Colorado, 1:100,000. U.S. Geological Survey, Miscellaneous Investigations Series, Map I-2400.

Hijmans. R.J. and J. van Etten. 2012. raster: Geographic data analysis and modeling. R package version 2.0-41. http://CRAN.R-project.org/package=raster

Liaw, A. and M. Wiener. 2002. Classification and Regression by randomForest. R News 2(3), 18-22.

O'Kane, S.L. 1988. Colorado's rare flora. Great Basin Naturalist 48:434-484.

NatureServe Explorer. 2012. An online encyclopedia of life [Web application]. Version 7.1. Arlington, VA. Available online: http://www.natureserve.org/explorer.

Pebesma, E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. R News 5 (2). http://cran.r-project.org/doc/Rnews/.

Phillips, S.J., M. Dudik, and R.E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Pages 655-662 in Proceedings of the 21st International Conference on Machine Learning. ACM Press, New York.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

Prasad, A.M., L.R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199.

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/ .

Rollins, R.C. 1983. Studies in the Cruciferae of western North America. Journal Arnold Arboretum 64: 491-510.

Rollins, R.C. 1984. Studies in the Cruciferae of western North America II. Contributions Gray Herbarium 214: 1-18.

Theobald, D.M., D.L. Stevens, Jr., D. White, N.S. Urquhart, A.R. Olsen, and J.B. Norman. 2007. Using GIS to generate spatially-balanced random survey designs for natural resource applications. Environmental Management 40:134-146.

Warren, D.L., R.E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33:607-611.

Warren, D.L., and S.N. Seifert. 2011. Environmental niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications 21:335-342.

APPENDIX A: LIST OF 1:24,000 GEOLOGIC QUADS DIGITIZED

maps available at the National Geologic Map Database: http://ngmdb.usgs.gov

Barnum, B.E. and Garrigues, R.S., 1980, **Geologic map and coal sections of the Cactus Reservoir quadrangle, Rio Blanco and Moffat Counties, Colorado**: U.S. Geological Survey, Miscellaneous Field Studies Map MF-1179, scale 1:24,000.

Barnum, B.E. and Hail, W.J., 1996, **Geologic map of Gillam Draw quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey, Miscellaneous Field Studies Map MF-2314, scale 1:24,000.

Cashion, W. B., 1969, **Geologic map of the Black Cabin Gulch quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Geologic Quadrangle 812.

Donnell, J. R., 1982, **Preliminary geologic map of the Sagebrush Hill Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 1398.

Donnell, J. R.; Hail, W. J., 1984, **Preliminary geologic map of the Calamity Ridge Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 1690.

Duncan, D. C., 1976, **Preliminary geologic map of Greasewood Gulch Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 755.

Duncan, D. C., 1976, **Preliminary geologic map of Jessup Gulch Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 756.

Duncan, D. C., 1976, **Preliminary geologic map of Rock School Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 757.

Duncan, D. C., 1976, **Preliminary geologic map of Square S Ranch Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 754.

Duncan, D. C., 1976, **Preliminary geologic map of Wolf Ridge Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 753.

Duncan, D. C., 1976, **Preliminary geologic map of Yankee Gulch Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 758.

Hail, W.J., Jr. and Barnum, B.E., 1993, **Geologic map of the Divide Creek quadrangle, Rio Blanco and Moffat counties, Colorado**: U.S. Geological Survey, Miscellaneous Field Studies Map MF-2232, scale 1:24,000.

Hail, W.J., Jr., 1973, **Geologic map of the Smizer Gulch quadrangle, Rio Blanco and Moffat Counties, Colorado**: U.S. Geological Survey, Geologic Quadrangle Map GQ-1131, scale 1:24,000.

Hail, W.J., Jr., 1974, **Geologic map of the Rough Gulch quadrangle, Rio Blanco and Moffat Counties, Colorado**: U.S. Geological Survey, Geologic Quadrangle Map GQ-1195, scale 1:24,000.

Hail, William James, 1984, **Geologic map of the Barcus Creek quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Geologic Quadrangle 1578.

Hail, William James, 1988, **Geologic map of the Barcus Creek SE quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Geologic Quadrangle 1613.

Johnson, R.C. and Smith, M.C., 1993, **Geologic map of the Philadelphia Creek quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey, Miscellaneous Field Studies Map MF-2216, scale 1:24,000.

O'Sullivan, R. B., 1974, **Preliminary geologic map of the Segar Mountain Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 570.

O'Sullivan, R. B., 1987, **Preliminary geologic map of the No Name Ridge Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 1989.

Pipiringos, G. N.; Johnson, R. C., 1975, **Preliminary geologic map of the Buckskin Point Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 651.

Pipiringos, G. N.; Johnson, R. C., 1976, **Preliminary geologic map and correlation diagram of the White River City Quadrangle, Rio Blanco County, Colorado**: U.S. Geological Survey Miscellaneous Field Studies Map 736.

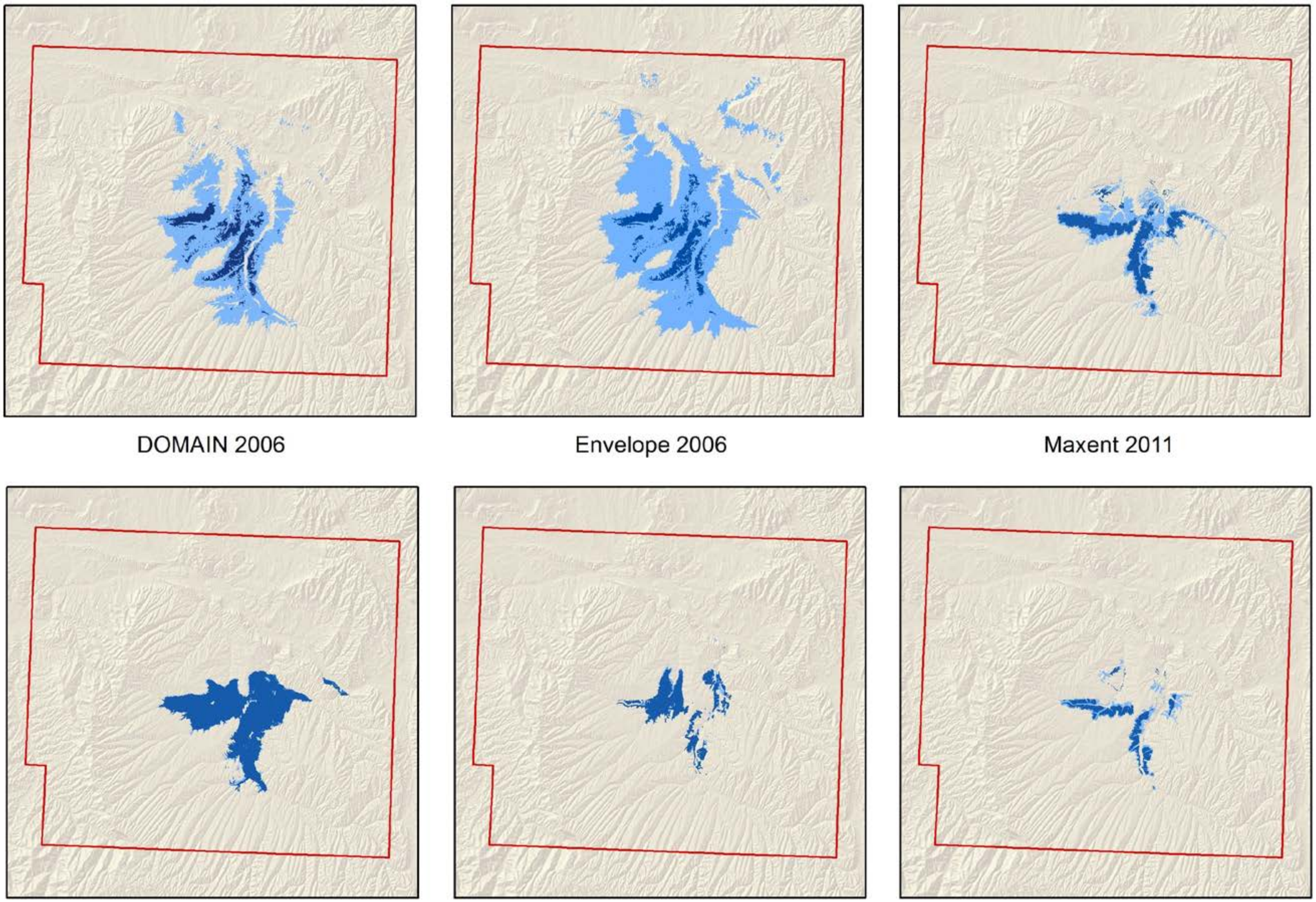
Pipiringos, G.N. and Rosenlund, G.C., 1977, **Preliminary geologic map of the Indian Valley quadrangle, Rio Blanco and Moffat Counties, Colorado**: U.S. Geological Survey, Miscellaneous Field Studies Map MF-836, scale 1:24,000.

Pipiringos, G.N. and Rosenlund, G.C., 1977, **Preliminary geologic map of the White Rock quadrangle, Rio Blanco and Moffat Counties, Colorad**o: U.S. Geological Survey, Miscellaneous Field Studies Map MF-837, scale 1:24,000.

APPENDIX B: SIDE-BY-SIDE COMPARISON OF MODELS

Models from 2006, 2011, and 2013 are shown for each species on the following pages. Areas of highest probability of suitable habitat are shown in dark blue, and areas of lesser probability are in light blue.

Physaria congesta models



DOMAIN 2006

Envelope 2006

Maxent 2011

Random forest - pseudo absence 2013

Random forest - true absence 2013

Maxent 2013

Physaria obcordata models