The Potential Distribution and Landscape Integrity

of

Frankenia jamesii (James' seaheath)

in Southeastern Colorado

photo by G. Doyle

prepared for:
Denver Botanic Garden

Karin Decker

Colorado Natural Heritage Program
Colorado State University
Campus Delivery 8002
254 General Services Building
Ft. Collins, Colorado 80523-8002

July, 2007

Knowledge to Go Places

Table of Contents

List of tables and figures

INTRODUCTION

Element distribution modeling

The process of developing a predictive model of the distribution of a particular species or ecosystem goes by a variety of different names and may involve several different techniques. All such modeling is based on the ecological principle that the presence of species and ecosystems (i.e., elements of biodiversity, or "elements") 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 element 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 this study, Element Distribution Modeling (EDM) 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.

This study used a classification and regression tree approach (Breiman et al. 1983) to investigate the potential distribution of *Frankenia jamesii*. The model was developed through a computerized procedure of binary recursive partitioning (Lewis 2000), wherein each group of element presence or absence points is successively split into two new groups, based on the values of independent (environmental) variables for those points. Modeling techniques are further discussed under "Methods" below.

It is important to regard these models as hypotheses intended to be field tested, and not as definitive maps of suitable habitat. A variety of life-history and biogeographic factors may preclude the presence of the target element in areas of predicted suitable habitat. Likewise, errors or lack of precision in modeling assumptions, input data, or procedures may incorrectly predict suitable habitat where none exists. In addition, users should be aware that the resolution of these distribution models is only as fine as the coarsest layer of input data (in this case 1 km-square cells). It is not appropriate to base land management decisions of 1-1000 m scale entirely on this analysis without additional field verification.

Study element

Frankenia jamesii Torr. ex Gray is a perennial shrub in the Frankeniaceae or Alkali-heath family. The species was named by Torrey in comparing a fragment collected by Edwin James, botanist on the Long expedition of 1820, with a specimen collected by Charles Wright during an 1849 expedition across the Rio Grande Valley to El Paso, Texas. Asa Gray (1873) described the species under Torrey's designation using specimens from the bluffs of the Arkansas near Pueblo, provided by E. L. Greene. *Frankenia jamesii* is the only member of this genus found in southeastern Colorado (Snow 1990). Its distribution in North America also includes southwestern Colorado in the Four Corners area, northern New Mexico, and western Texas. *Frankenia jamesii* is ranked G4 (apparently secure) by NatureServe, and is not ranked in Colorado. The species has been reported from Fremont, El Paso, Pueblo, Otero, Kiowa, Bent, and Las Animas Counties in southeastern Colorado, and Montezuma County in southwestern Colorado. The species is described as occurring on gypsiferous soils and alkaline shales (Weber and Wittmann 2001, Holmgren 2005).

Figure 1. ***Frankenia jamesii***

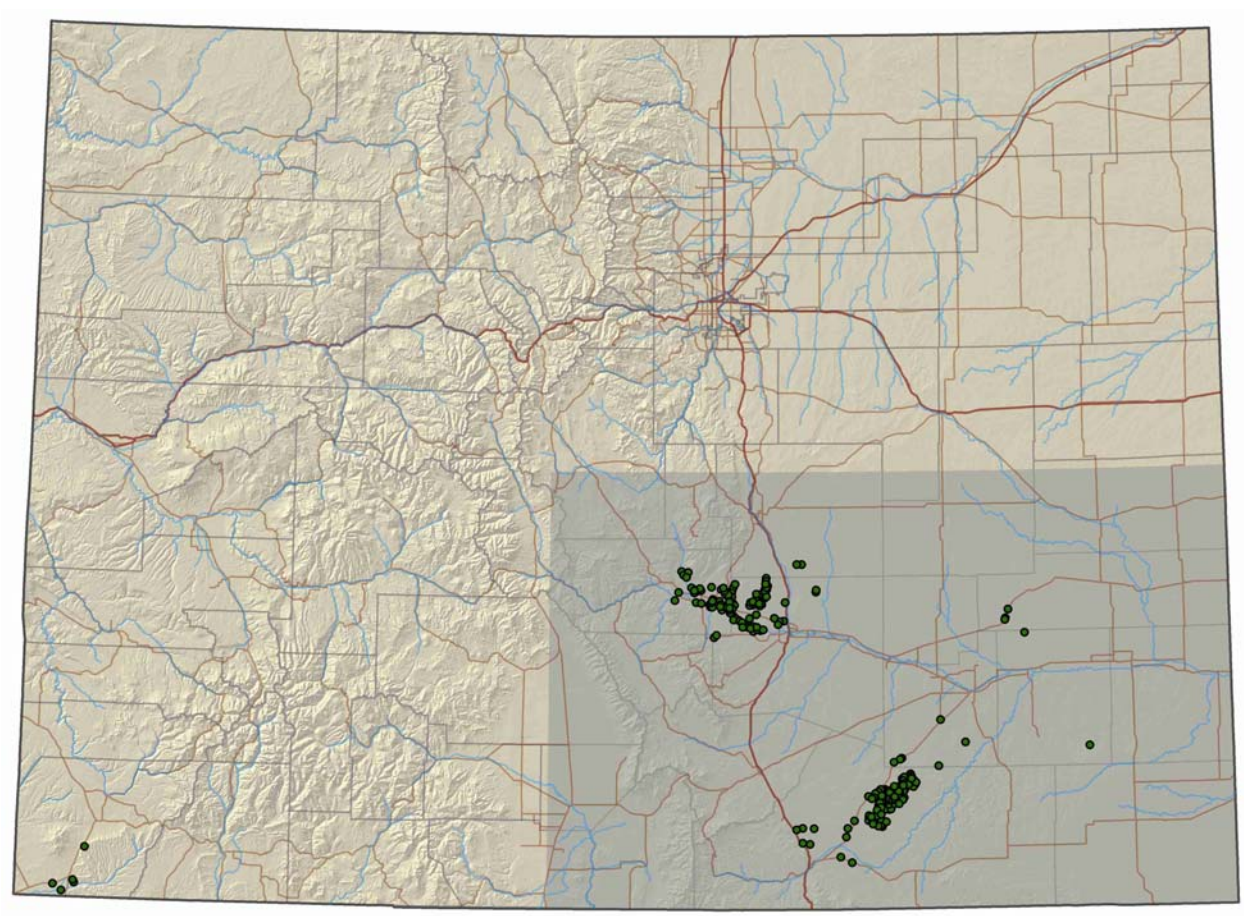
photo by S. Spackman Panjabi

METHODS

Input data

The study area was restricted to the southeastern quarter of Colorado (Figure 2). Although *Frankenia jamesii* also occurs in the southwestern corner of the state, there was insufficient data to allow reasonable modeling of this area. Models were constructed with data from documented locations of the target species using element occurrence records from the Colorado Natural Heritage Program database (CNHP 2007) and herbarium records from COLO and CSU. No historic or known extirpated records were used. Element occurrence records included 82 plant records with *F. jamesii* listed as an associated species, and 9 community records for associations including *F. jamesi*i. Element occurrence polygons were converted to point locations. Up to 10 randomly placed points per polygon were generated in order to more accurately represent the extent of the polygon. Point locations were added to this dataset from 10 herbarium records with good location descriptions. Finally, 240 points were randomly selected such that no point was within 1000m of another point. From these positive model points, 60 were withheld from the modeling dataset for later use in model validation. This resulted in a modeling dataset of 180 positive points.

Figure 2: Study area and presence points.



Absence data were generated for the modeling process by two separate methods. The first absence dataset was compiled from point locations of plant element occurrences that did not list *Frankenia jamesii* as an associated species. The assumption was that these "pseudo-absence" points represent locations where there is a reasonable assumption that the target species was in fact absent, since it would typically have been reported by CNHP staff during survey. From the compiled dataset of 625 locations, a sample of 303 points was selected randomly such that no point was within 1000 m of another, or within 2000 m of a known location of the target species (Figure 2). Again, no historical records were used. A second pseudo-absence dataset was produced by randomly generating 125 points in each quadrant of the study area, and removing any that fell within 2000m of a known presence point. This resulted in a random pseudo-absence dataset of 491 points. One hundred of the absence points were withheld from each modeling dataset for validation.

Environmental attributes for both presence and absence points were derived from digital raster data in ArcGIS 9.2 (ESRI 2006). Datasets were processed to a common projection, clipped to the study area, and resampled as necessary to a 30 m cell size. Environmental data used and sources are listed in Table 1.

Classification and regression tree modeling

Classification and regression analyses use a variety of algorithms for predicting continuous or categorical variables from a set of continuous or categorical effect variables (Breiman et al. 1983). 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.). In this study, I used a simple binary classification-type analysis predicting the presence or absence of a species according to the values of various environmental factors. At each iteration, the recursive partitioning 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.

An important issue in the use of classification and regression tree analyses is deciding when to stop splitting. The recursive partitioning process can continue to split datasets until all environmental variables have been accounted for and each terminal node is composed of strictly one class or the other (i.e. overfitting). Real-world data typically contains random error or noise that may result in splits which are not ecologically meaningful. Overfit models, while perfectly predicting the distribution of locations used in the model, may be less accurate in predicting independent validation points. The general approach to "pruning" the classification tree is to stop generating new split nodes at a point when subsequent splits give only a small overall improvement of the level of prediction.

Classification and regression tree analysis was implemented in the open-source software program R 2.4.1 (R Development Core Team 2006), using the rpart recursive partitioning package (Therneau and Atkinson 2007). Pruning was accomplished by using the cross-validation complexity parameter (cp) generated by the rpart routine. The 1-SE rule (Breiman et al. 1983) was applied by choosing the value of cp associated with the simplest tree that minimized cross- validation error, and pruning the tree to that level (Atkinson and Therneau 2000).

Table 1: Environmental variables used in modeling.

Continuous Variables	**Units**	**Source**
Elevation	m	USGS 30m Digital Elevation Model (DEM) for Colorado
Slope	degrees	Derived from DEM
Total annual precipitation	cm	Daymet - Climatological summaries for the coterminous United States 1980-1997 http://www.daymet.org/ (1km)
Precipitation frequency (proportion of wet days)	proportion	Daymet
Monthly precipitation (12 separate months)	cm	Daymet
March minimum air temperature	°C	Daymet
April minimum air temperature	°C	Daymet
May minimum air temperature	°C	Daymet
Number of frost days	days	Daymet
Categorical Variables	**Values**	**Source**
Aspect	N, NE, E, SE, S, SW, W, NW, Flat	Derived from DEM
Surface Geology	Various	Colorado State Geologic Survey. 1995. The Digital Geologic Map of Colorado in ARC/INFO Format. From Tweto, O. 1979. Geologic Map of Colorado, with details of Niobrara formation added from 1 x 2 minute quads.
Soil type	Various, see Appendix	USDA Soil Conservation Service. 1994. General Soil Associations (STATSGO) for Colorado.
Vegetation type	Various	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.

Model validation

Several models were tested with an independent validation dataset of points randomly withheld from the original datasets. Validation points (both presence and absence) were overlaid on the distribution maps to determine the number of correct identifications, the number of false positives, and the number of false negatives. Precision and accuracy of the resulting classification were calculated as shown below:

CP Correct Positive FP False Positive CN Correct Negative FN False Negative	Precision: The proportion of predicted positive cases that were correct. = CP / (CP + FP) Accuracy: The proportion of the total number of predictions that were correct. = (CP + CN) / (CP + CN + FP + FN)

Note that precision and accuracy as calculated above are sensitive to the relative proportion of presence and absence points. A low proportion of presence points makes this a less useful measure of model success.

Landscape integrity analysis

A Landscape Integrity GIS dataset for Colorado developed by the Colorado Natural Heritage Program (CNHP 2006), was used to score known *Frankenia jamesii* locations in Colorado. This dataset represents the cumulative impacts from oil and gas wells, surface mining, urban development, agriculture, and roads that threaten the viability of ecological systems within the state of Colorado as of 2006. The model is based on distance decay functions of modified s-curves for each input threat. By adjusting the shift and spread of the curve, the rate of decay (i.e. the decreasing impact of a particular threat), and the weight (severity) can be tailored to specific threats. The curves created are asymptotic at both ends and were truncated as shown below. The individual threat layers were combined into a single landscape integrity dataset (Figure 3).

Table 2. Components of Landscape Integrity GIS dataset.

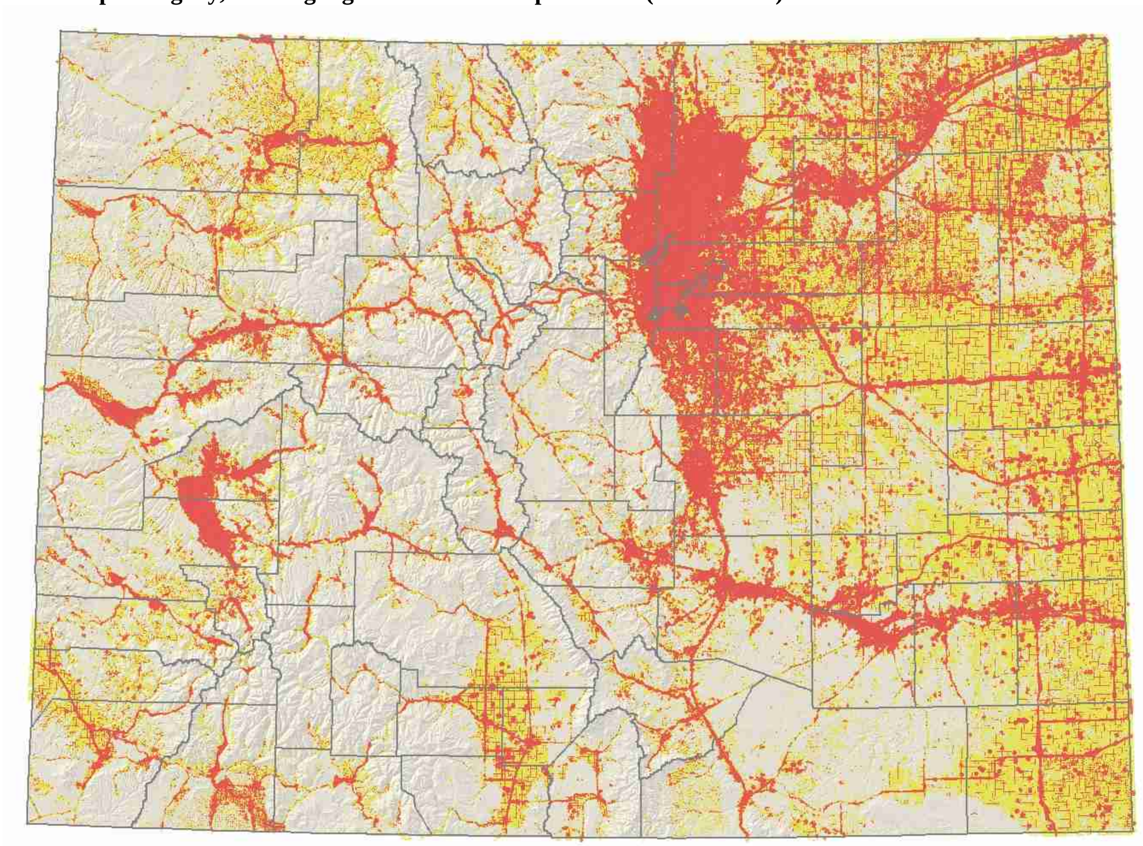
Threat type	Weight	Distance decay function type
All development	500	gradual
Surface Mines	500	moderate
Roads primary & secondary	500	moderate
Oil & gas wells	400	moderate-abrupt
Agriculture	300	moderate
Roads - local & primitive	200	abrupt

Where:

Decay Function	Cut off Distance	Equation
Abrupt	250m	(1 / (1 + Exp(((Distance / 100) - 1) * 5))) * Weight
Moderate-Abrupt	600m	(1 / (1 + Exp(((Distance / 100) - 2.5) * 2))) * Weight
Moderate	1250m	(1 / (1 + Exp((Distance / 100) - 5))) * Weight
Gradual	2000m	(1 / (1 + Exp(((Distance / 100) - 10) * 0.5))) * Weight

Mean landscape integrity scores by county were produced by the Zonal statistics routine in ArcGIS.

Figure 3: Landscape Integrity, showing high and medium impact areas (CNHP 2006).



RESULTS

Predicted distribution models

The first models generated with each type of absence data both had soil type as the first splitting factor. The EO-absence model (Soils-EO) also included October precipitation (Figure 4a), while the Random-absence model (Soils-Rand) included elevation and January precipitation (Figure 4b). The Random-absence model is more restricted in scope than the EO-absence model (Figure 5)

Validation results are shown in Tables 3 and 4.

Table 3: Model Soils-EO validation.

Soils-EO	**Model present**	**Model absent**
Known present	55 (CP)	5 (FN)
Known absent	6 (FP)	94 (CN)
% correct positives: 91.6% % correct negatives: 94%	**Precision = .90** **Accuracy = .93**	

Overall model points correctly classified: 91.5%

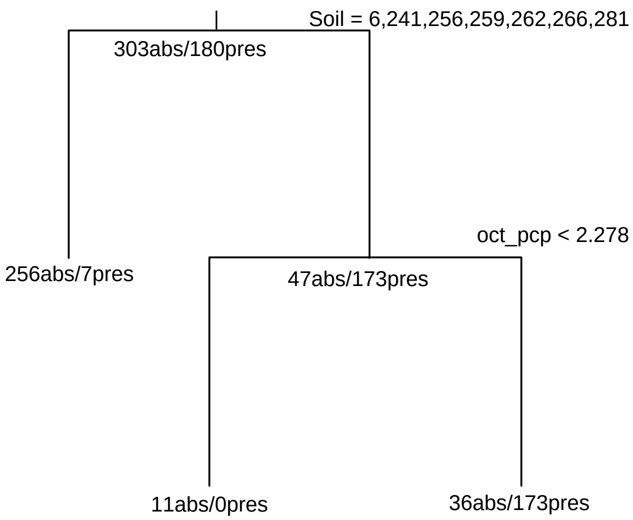
Table 4: Model Soils-Rand validation.

Soils-Rand	**Model present**	**Model absent**
Known present	54 (CP)	6 (FN)
Known absent	6 (FP)	94 (CN)
% correct positives: 90% % correct negatives: 94%	**Precision = .90** **Accuracy = .925**	

Overall model points correctly classified: 95%

Figure 4: Classification trees for Model Soils-EO (a) and Soils-Rand (b).

(a)



(b)

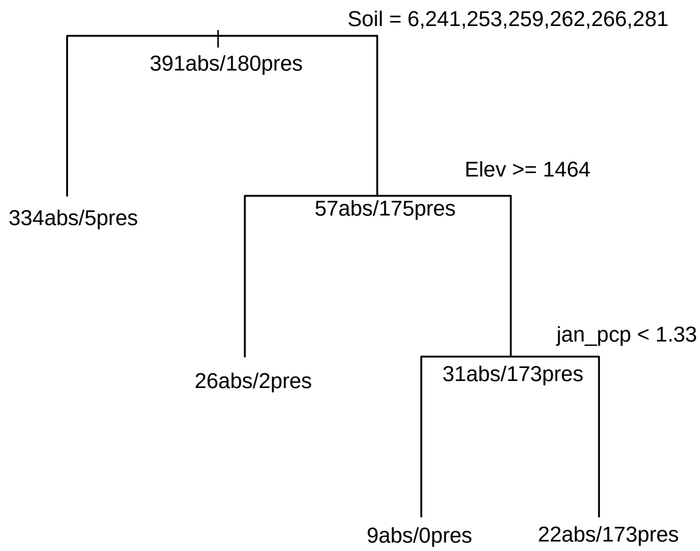
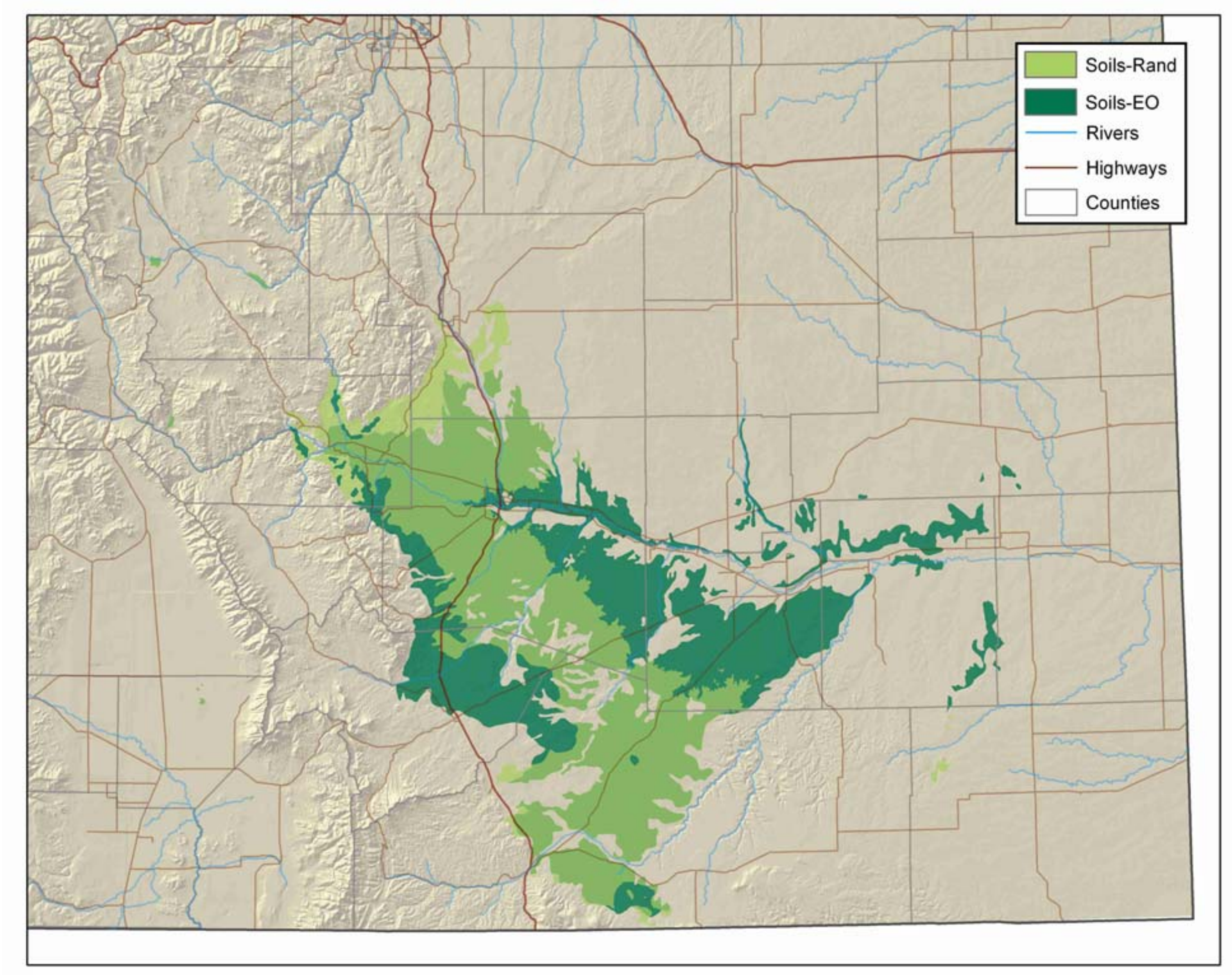


Figure 5: Map of potential distribution of *Frankenia jamesii* derived from Models Soils-EO and Soils-Rand.



In order to investigate the effect of surface geology on the model, soil types were combined into Great Groups, reducing the number of types in the analysis. Subsequently, models with both absence datasets selected geology as the primary splitting factor, in combination with various monthly precipitation variables. Model Geol-EO, produced using EO-absence data, selected geology and June precipitation as splitting factors (Figure 6a), and produced the most extensive predicted suitable habitat of any model (Figure 7). Model Geol-Rand, produced using the random-absence data, was the most complex model, selecting geology, June, December, August and May precipitation, with two terminal nodes retained in the model (Figure 6b). Validation results are shown in Tables 5 & 6.

Table 5: Model Geol-EO validation .

Geol-EO	**Model present**	**Model absent**
Known present	57 (CP)	3 (FN)
Known absent	8 (FP)	92(CN)
% correct positives: 95% % correct negatives: 92%	**Precision = .877** **Accuracy = .93**	

Overall model points correctly classified: 89.6%

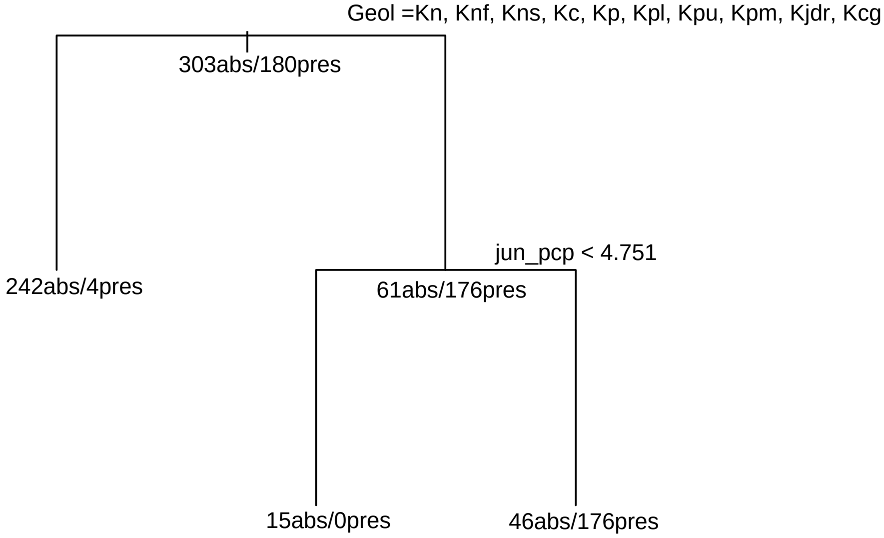
Table 6: Model Geol-Rand validation .

Geol-EO	**Model present**	**Model absent**
Known present	54 (CP)	6 (FN)
Known absent	7 (FP)	93(CN)
% correct positives: 90% % correct negatives: 93%	**Precision = .885** **Accuracy = .92**	

Overall model points correctly classified: 95.4%

Figure 6: Classification trees for Model Geol-EO (a) and Geol-Rand (b).

(a)



(b)

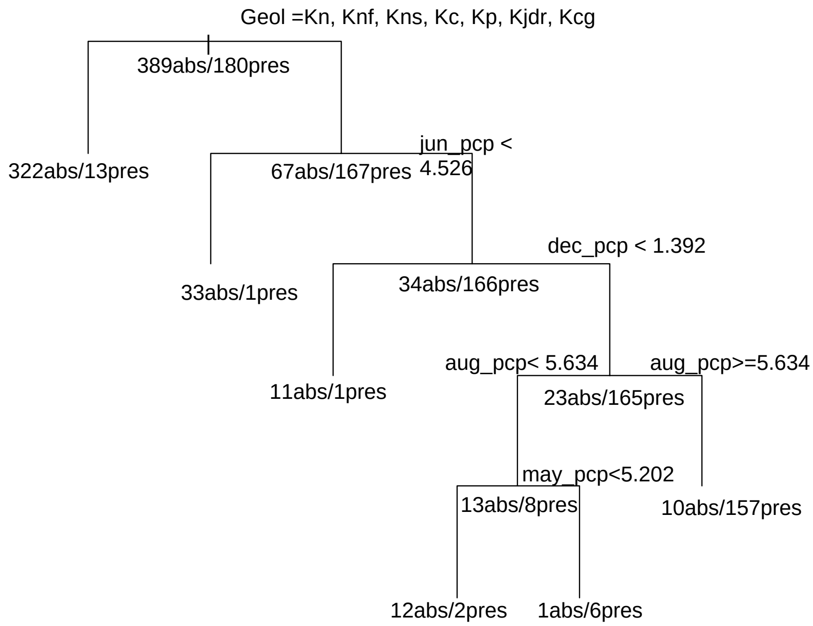
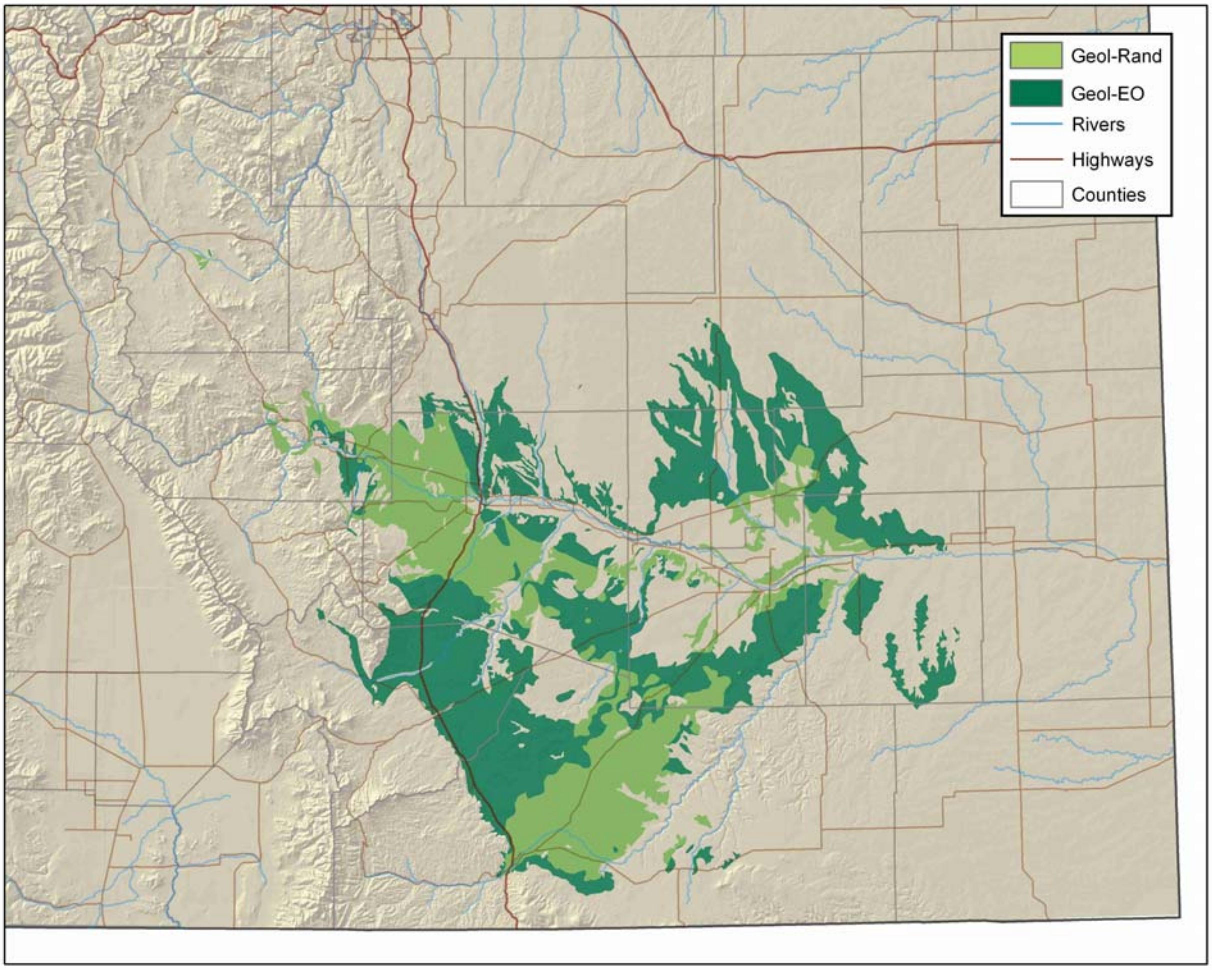


Figure 7: Map of potential distribution of *Frankenia jamesii* derived from Models Geol-1 & Geol-2.

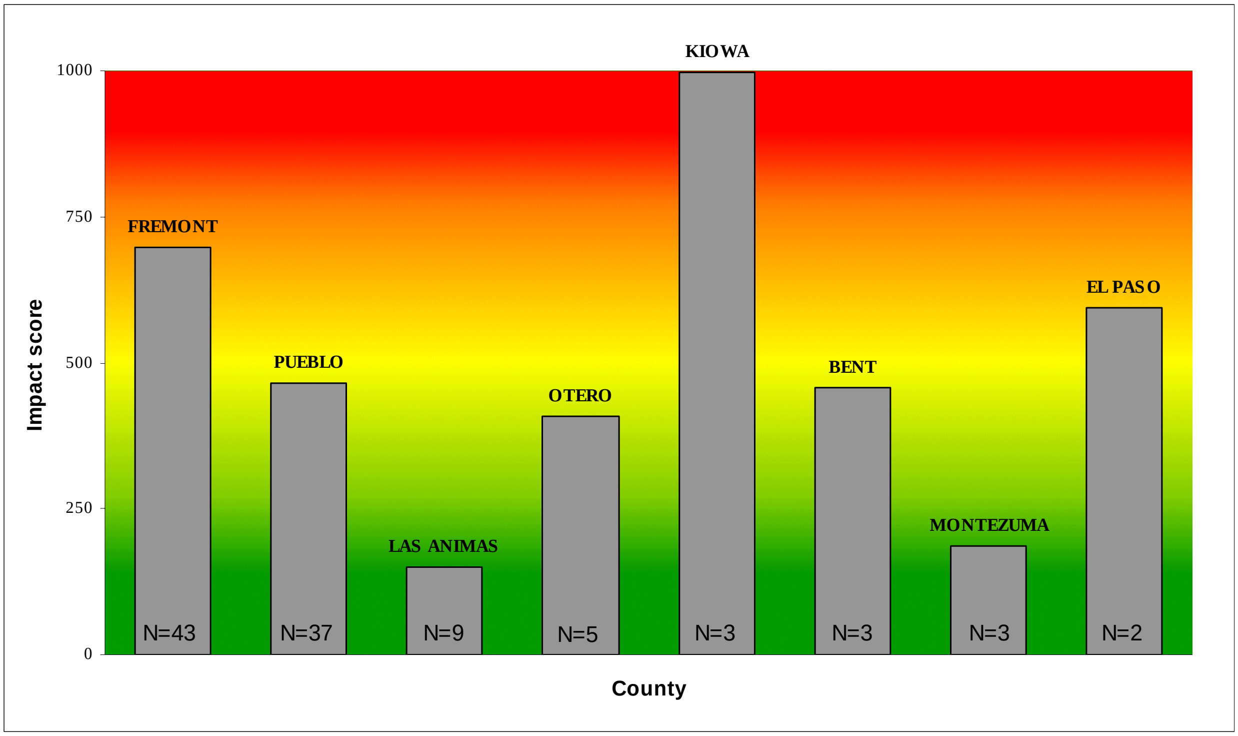


Landscape Integrity Analysis

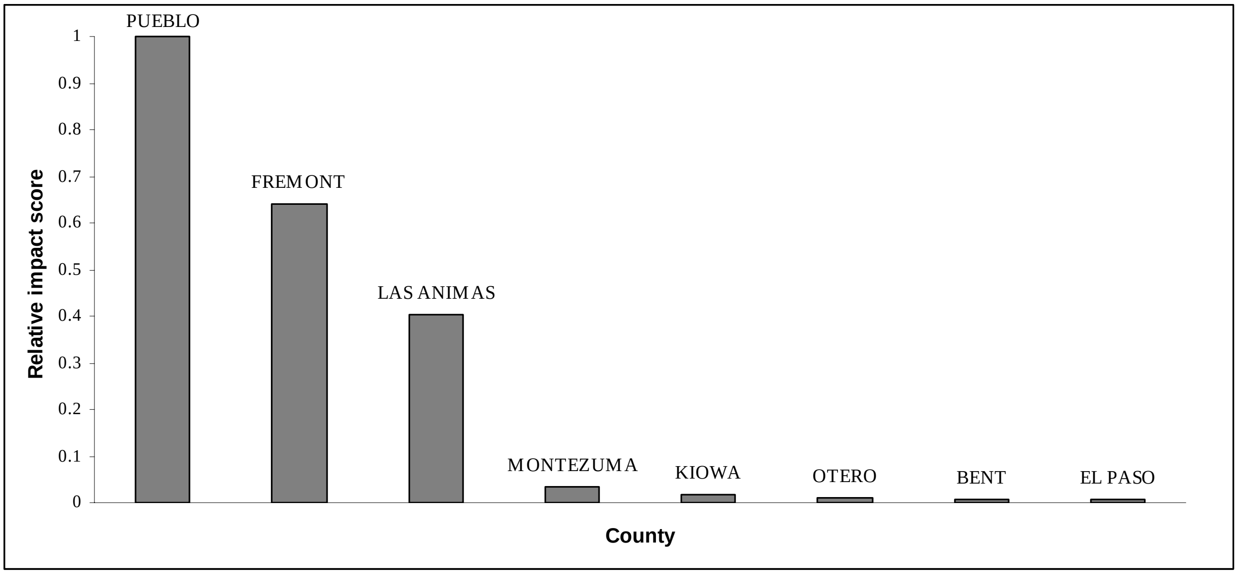
The mean Landscape Integrity score for *Frankenia jamesii* occurrences in Colorado is 357. This level was considered Moderate Impact in Decker et al. (2007). Impact scores by county are shown in Figure 8a, together with the number of occurrences from that county. Counties are arranged in order of number of occurrences. Area-weighted, relative impacts scores by county are shown in Figure 8b. Fremont and Pueblo Counties have the highest numbers of document occurrences, and these are concentrated in areas of higher impact.

Figure 8: Landscape integrity impact scores by county.

(a)



(b)



DISCUSSION

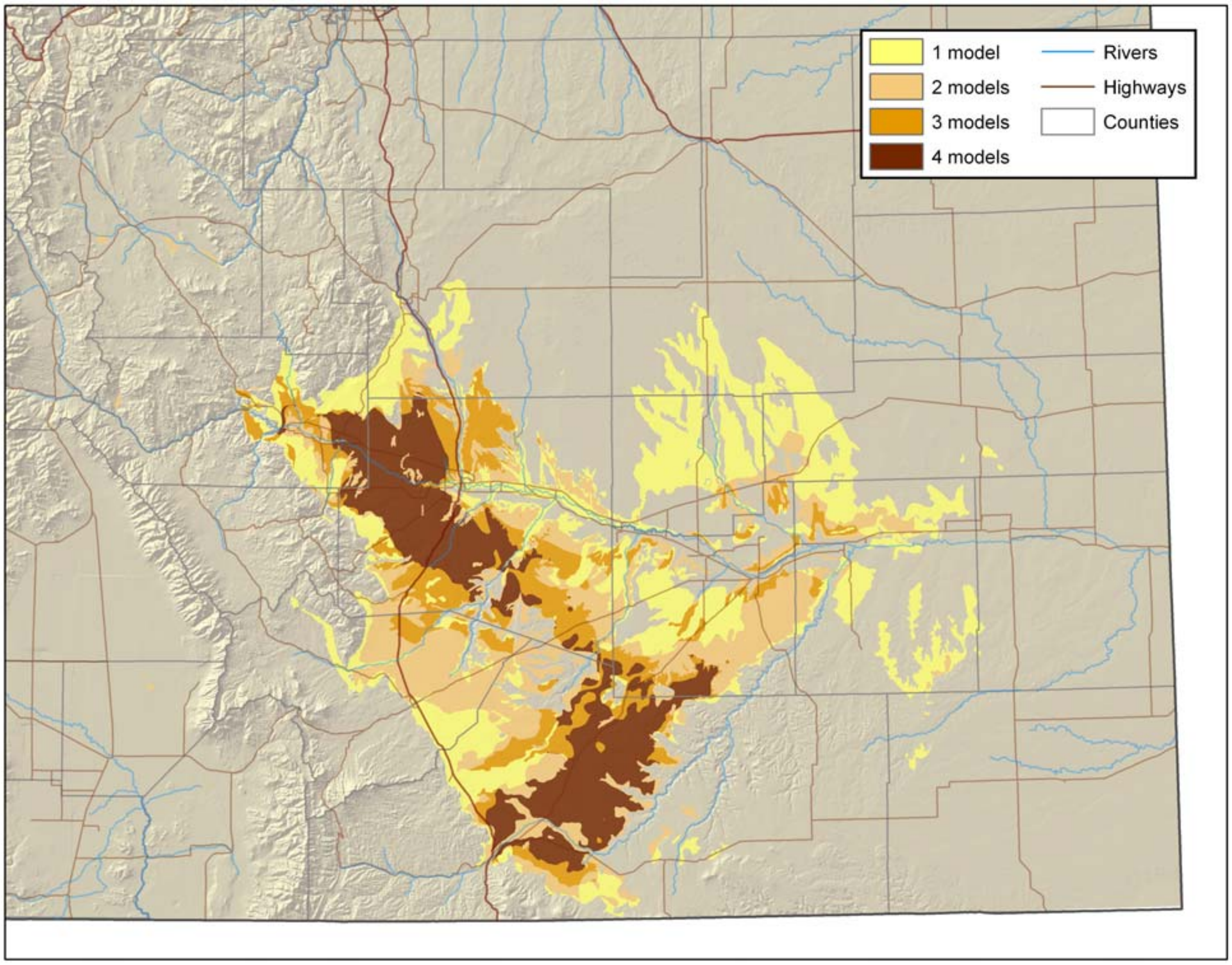
The Arkansas River basin in southeastern Colorado is primarily underlain by the Niobrara and adjacent cretaceous age formations. Substantial areas of the basin are covered by Quaternary alluvial material, leaving the Cretaceous formations more exposed in a broadly L-shaped feature extending southeast from Colorado Springs to the state border with New Mexico, and bending up to the northeast toward the vicinity of Cheyenne Wells. The predicted range of *Frankenia jamesii* roughly follows the mapping of the Niobrara formation in southeastern Colorado. Primary formations are the Niobrara (Kn, Kns, and Knf), Carlile Shale, Greenhorn Limestone, and Graneros Shale (Kcg) and the Pierre Shale (Kp, Kpu, Kpm, Kpl). The most common soils corresponding to *F. jamesii* occurrences include the Manzanola, Limon, Kim, and Midway series, which are all slightly to strongly alkaline, well drained soils derived from shale, sandstone, or clay.

All four models have similar validation results, with reasonably high precision and accuracy in predicting presence and absence points, although they are clearly different in extent. Because neither soils nor geology are mapped at very fine scales throughout the range of *Frankenia jamesii* in southeastern Colorado, it is difficult to recommend one particular model over another. Instead, the four models were combined to produce a single layer indicating where multiple models are in agreement (Figure 9), with the interpretation that areas of highest overlap are areas of the highest probability of suitable habitat.

Landscape integrity scores indicate that in certain areas of southeastern Colorado *Frankenia jamesii* may be more vulnerable to anthropogenic impacts. In particular, populations in Pueblo county have a higher relative impact score than can be strictly accounted for by the prevalence of occurrences in that county. Scores from Kiowa County are heavily influenced by the fact that the few documented occurrences are all roadside sites.

The results of this study indicate that there may be substantial tracts of potentially suitable habitat for *Frankenia jamesia* in southeastern Colorado. The models do not address population density, and plants may be sparse to scattered within suitable habitat. These models should be evaluated by field survey, and the results used to further refine the predictions. It is not appropriate to base land management of conservation planning decisions on this analysis without additional field verification.

Figure 9. Combined model of predicted distribution for *Frankenia jamesii.*



REFERENCES

Atkinson, E.J. and T.M. Therneau. 2000. An Introduction to Recursive Partitioning Using the RPART Routines (short version of Therneau and Atkinson 1997). Mayo Clinic, Section of Statistics, Rochester MN. http://ndc.mayo.edu/mayo/research/biostat/upload/rpartmini.pdf

Breiman L., J.H. Friedman, R.A. Olshen, and C.J Stone. 1983. *Classification and Regression Trees*. Wadsworth, Belmont, Ca,

Colorado Natural Heritage Program (CNHP). 2007. Biodiversity Tracking and Conservation System (BIOTICS). Colorado Natural Heritage Program, Colorado State University, Fort Collins, CO.

Colorado Natural Heritage Program (CNHP). 2006. Landscape integrity in Colorado. GIS dataset. Colorado Natural Heritage Program, Colorado State University, Fort Collins, CO.

Decker, K, G. Doyle, C. Gaughan, R. Rondeau, M. Fink, and A. Lavender. 2007. A Biodiversity Scorecard for Colorado. Colorado Natural Heritage Program, Colorado State University, Fort Collins, CO.

ESRI. 2006. ArcGIS 9.2. ESRI, Redlands, CA.

Gray, A. 1873. Characters of New Genera and Species of Plants: *Frankenia jamesii.* Proceedings of the American Academy of Arts and Sciences. 8:623.

Holmgren, N.H. 2005. Frankeniaceae. *In* Intermountain Flora: Vascular Plants of the Intermountain West, USA, Volume 2, Part B. Holmgren, N.H., P.K. Holmgren, and A. Cronquist. New York Botanical Garden. New York.

Lewis, R.J. 2000. An Introduction to Classification and Regression Tree (CART) Analysis. Presentation to the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California. Available online at: http://www.saem.org/download/lewis1.pdf

NatureServe. 2005. NatureServe Explorer: An online encyclopedia of life [web application]. Version 4.5. NatureServe, Arlington, Virginia. Available at: http://www.natureserve.org/explorer

R Development Core Team. 2006. 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

Snow, N. 1990. The distribution of *Frankenia jamesii* Torr. ex Gray (Frankeniaceae) in the Great Plains. Prairie Naturalist 22(3):211-214.

Therneau, T.M. and E.J. Atkinson. 1997. An introduction to recursive partitioning using the RPART routine. Technical Report 61. Mayo Clinic, Section of Statistics, Rochester MN. Available online at http://www.mayo.edu/hsr/techrpt/61.pdf

Therneau, T.M. and E.J. Atkinson. R port by Brian Ripley <ripley@stats.ox.ac.uk>. 2007. rpart: Recursive Partitioning. R package version 3.1-36. S-PLUS 6.x original at http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm

Tweto, O. 1979. Geologic Map of Colorado.

Weber, W.A. and R.C. Wittmann. 2001. *Colorado Flora: Eastern Slope*. University Press of Colorado, Niwot, Colorado.