POTENTIAL DISTRIBUTION MODELING OF SENSITIVE AND RARE PLANT SPECIES IN SHOSHONE NATIONAL FOREST

Prepared by

The Wyoming Natural Diversity Database

University of Wyoming

Laramie, Wyoming

December 18, 2016

Cooperative Agreement No. 11-CS-11021400-008, Mod. 04

CONTENTS

ABSTRACT

We produced predictive distribution models for 38 plant species that occur on or adjacent to the SNF, using a common algorithm and methods. These Wyoming models can be used to guide surveys looking for new populations of each species, and can also be used as a "coarse filter" for evaluating the potential impacts of management activities or development on the species. As with all models, uncertainty exists, and their use on SNF might be enhanced by environmental data layers unique to the Forest. These distribution models should not be used as a substitute for field surveys.

INTRODUCTION

The Shoshone National Forest (SNF) manages for Sensitive[1] species and species of local concern (SOLC[2]). Sensitive species in particular receive special emphasis in planning and management activities to assure their conservation. On the SNF, this includes 24 Sensitive plant species as well as 23 plant species with SOLC designation.

Effective management of priority plant species requires an understanding of where a species is found on the landscape. Existing data provide an indication of where priority species are currently known to occur on the SNF, but data coverage is invariably uneven. Species distribution modeling can be used to evaluate the potential for the occurrence of a species in unsurveyed areas, and has become a common tool for informing management of priority species in Wyoming[3-7], and elsewhere. Distribution models relate the occurrence of a species to important environmental gradients using statistical methods, and then extrapolate these statistical relationships using Geographic Information Systems (GIS), to make predictions about the likelihood of occurrence for the species in unsurveyed areas. WYNDD has produced models for other rare plant species in the state[3, 6, 8], but models for most of the SNF Sensitive and SOLC species have not been generated prior to this project.

Many of the predictor datasets and methods used to generate models as part of this project were the same as those used for a previous project aimed at modeling the distribution of rare plants on BLM lands in Wyoming[8]. As such, we describe these datasets and methods more generally in this report, and refer readers who need additional information to the previous report[8] for more detail. In cases where more detailed information was deemed necessary to include in this report to effectively describe the datasets or methods used, text has been included, in some cases verbatim, from the BLM plant modeling project report. The majority of the information related to species-specific model results can be found in Appendix 2, which contains summary statistics, thumbnails of model output, and other model-specific details.

METHODS

OVERVIEW

Plant species distribution records (hereafter referred to as occurrence records) have been compiled in a database managed by the Wyoming Natural Diversity Database (WYNDD; University of Wyoming). For this project, occurrence records for 38 species were exported from the database, along with a background, pseudo-absence dataset generated from the occurrence records of other plant species of concern in Wyoming, from the same database. GIS layers representing environmental gradients relating to climate, soils, hydrology, topography, and vegetation were assembled to use as predictor variables. Predictive distribution models were then generated using Random Forest[9], an extension of Classification and Regression Trees (CART) that uses an iterative and stochastic process to produce useful models from small training datasets. The resulting statistical models were used to produce GIS layers showing the predicted relative probability of occurrence for the candidate species set. Each step in the process is described in greater detail on the following pages.

Species Targets

The 24 Sensitive plant species and 23 plant species with SOLC designation on the SNF formed the initial set of candidate species for modeling. About 60 additional Wyoming plant species of concern or species of potential concern (WY-SOC) that are present on SNF were also considered as candidates for modeling[10]. Species in these three groups (Sensitive, SOLC, WY-SOC) were evaluated as prospective modeling targets by Michael Kirkpatrick, Forest Botanist in an initial scoring table that also included factors such as whether the species has known management concerns and whether it is an endemic. He proposed species from all three groups, with WYNDD input, and culled some of the species if there was another species target of very similar habitat requirements. SNF harbors some priority species only known from 1-2 occurrences; these species were eliminated from consideration as this was too few records to produce useful models.

Thirty-eight plant species were ultimately identified for modeling, including 14 Sensitive species, 8 SOLC, and 22 Wyoming species of concern (Table 1). They include species that are state endemics: species for which the Wyoming distribution encompasses worldwide distribution. They also include species that are regional endemics (geographically restricted, though with distribution extending beyond state borders), disjuncts (populations removed by about 300+ miles from the main, contiguous portion of the continental range), peripherals (at the perimeter of contiguous distribution), and species that are sparsely distributed throughout their range. There was some overlap between the three SNF plant lists and the previous set of species modelled for BLM in Wyoming[3, 8], and there was no adjustment to targets if a species had previously been addressed in the prior study. Specifically, eight species that were included as modeling candidates for the SNF had previously been modeled as part of a project for the Wyoming BLM; these species were not eliminated as potential modeling candidates since new occurrence or predictor data may have become available that could have produced improved models. Further, the non-deterministic nature of the Random Forest algorithm and the subsampling routine we employed in each project (explained below, in the *Model Generation, Validation, and Display* section) means that slightly different models could result from successive but identical model generation processes, even with identical input data. Producing another set of models for these eight species therefore could provide some insight into the potential variability in model output with these methods, with very little additional investment of time.

Presence Data Collection and Processing

WYNDD maintains plant occurrence data with supporting bodies of species information, taxonomic information, and reference information. The occurrence data are a synthesis of known distribution of Wyoming plant species of concern, including all TES species and other rare, native species whose viability is in question in the state. The WYNDD observation database integrates available documentation including the specimen data of RM and all other studies to produce a comprehensive and cohesive picture of each species' known distribution.

Systematic floristic surveys have been conducted on SNF through the Rocky Mountain Herbarium (RM) that have significantly expanded our understanding of the SNF flora[11-14]. Major contributions include decades of investigation by Erwin Evert[15] and Richard Scott[16]. Substantial floristic documentation was built into research natural area establishment reports prepared by WYNDD[17-25]. WYNDD also conducted targeted gap-filling studies of candidate sensitive species on the Forest[26]. Finally, of the 38 species, nine

have been the subject of one or more systematic survey projects, mostly by WYNDD, expanding and synthesizing distribution syntheses, as stored in WYNDD's database, and compilation of status-pertinent information in these status reports [27-39]. Some of the systematic survey projects were for BLM so the level of work on national forest was wanting or incomplete. Each survey project sought to expand known distribution using some combination of map interpretation, review of environmental information, review of existing distribution data, field reconnaissance, and extensive field work. As such, they have relatively robust presence data. There have also been segments of SNF targeted for gap-filling surveys in areas of high sensitive species concentrations (e.g., on the North Fork of the Shoshone River) insofar as management plans needed more robust data[40]. This resulted in a wide array of data documentation levels among the target species. This body of information is represented in WYNDD databases and in a one-time series of species assessment documents produced by WYNDD botanists and others to compile existing species' data throughout the USFS Rocky Mountain Region, including 36 species assessments that address 36 of the SNF sensitive and rare plant species (posted on http://www.fs.usda.gov/detail/r2/landmanagement/?cid=stelprdb5177128).

WYNDD's Biotics database was queried for all occurrence data used in building models. The database has a system to flag dubious records, and a secondary review step was built in to double-check the data exports for dubious records and remove them or any others that had unresolved identity questions (e.g., those specimens displaying features intermediate between two species, and of possible hybrid origin). Since species may no longer be extant at historical sites, relating historical records to the environmental gradients might not produce a model that accurately predicts current distribution. Thus, records representing observations from before 1970 were excluded. Likewise, occurrence records with a mapping buffer representing mapping uncertainty of greater (i.e., worse) than 1200 meters were also excluded. This distance is commonly applied to the many plant records that have no greater location detail than section. Excluding less precise records reduces the possibility that a poorly-mapped point location will introduce noise into modeling, by falsely indicating presence in an unsuitable setting.

Although species distribution modeling is typically based solely on points representing documented observations for the target species, plant records from WYNDD's database instead comprise two basic and distinct types of shapes: points and polygons, both of which are referred to as "source features." All records that have collection data as their most detailed source of information are mapped as points, with an uncertainty buffer applied, resulting in a circular feature. By contrast, almost all records that have survey data as their most detailed source of information are mapped as polygons of occupied habitat, created either as boundaries drawn on field maps and later digitized, or as waypoints collected in the field and subsequently used to draw polygons in GIS. Records that have survey data as their most detailed source of information are usually mapped as polygons, unless the entire source feature fits within an area of 20 m radius.

These two different source feature types required different processing methods in order to generate modeling points and relate the records to specific environmental gradient values (e.g., Figure 1). For buffered presence points, the centroid of each circular feature was used, as this minimized the potential spatial error in a point's location. For mapped polygons of occupied habitat, a grid of points was generated at 30 m spacing within each polygon, aligned with the centroids of the 30 m raster cells used to represent environmental predictors. By creating multiple, gridded points to represent each habitat polygon rather than using a single polygon centroid, more information was available from each polygon. Although using all gridded points in a model would represent a form of pseudoreplication[41], iterative

resampling (explained further in the "Model Generation" section, below) allowed the use of more of the information available in the gridded points while preventing pseudoreplication issues.

By these methods, approximately 100,000 presence points were generated, representing about 1,750 mapped occurrence locations for the target species. The number of mapped occurrences available by species varied dramatically, from 225 occurrence locations for Beaver Rim Phlox to just 9 for Twinleaf cinquefoil (Table 1), with a median of 27. Useful species distribution models have been generated by other researchers with as few as 5-50 presence points[42-49], though there appears to be some consensus that having greater than 50 presence points results in more robust models[42]. Twelve of the 38 modeled species had 50 or more presence points; 20 species had fewer than 30 presence points, a lower limit that is sometimes suggested as a minimum (see Franklin and Miller[42] for a thorough discussion of sample size considerations).
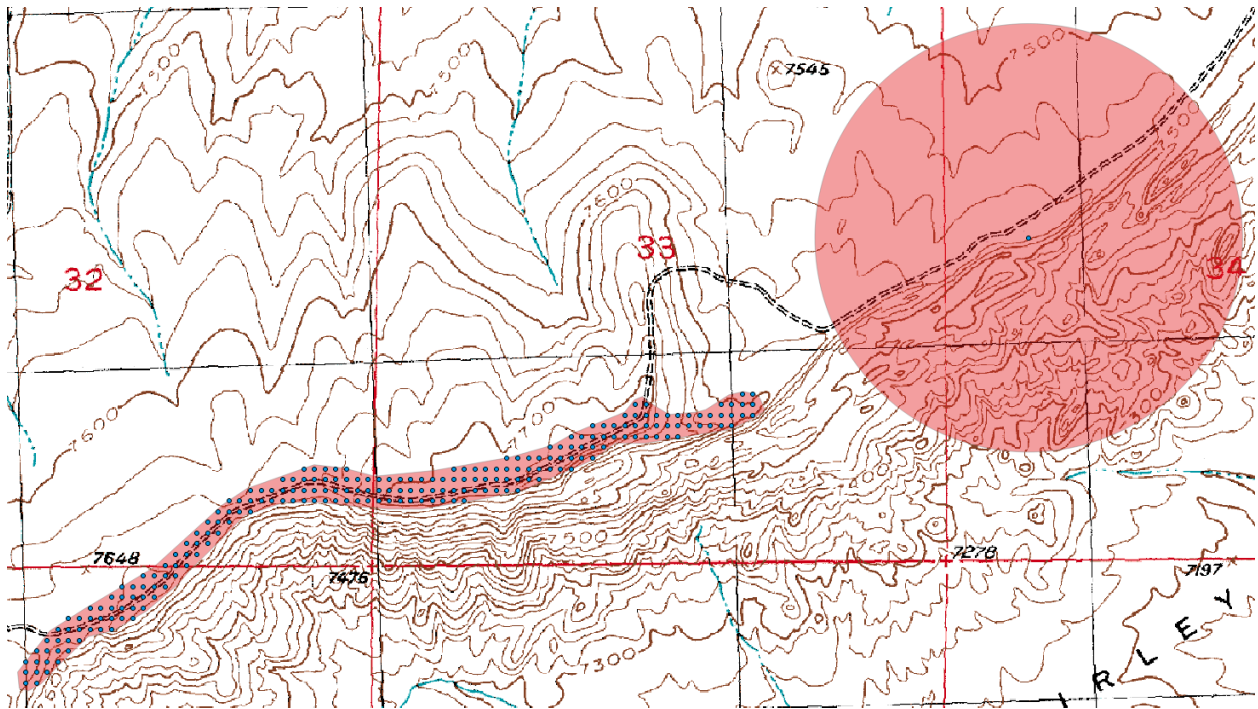

Figure 1. Polygon source features representing occupied habitat, as in the case of the elongate shape represented by a red polygon in the lower left, were sub-sampled using gridded points spaced at 30 m, shown here as blue dots. Source features representing buffered points were sub-sampled using a single centroid for the circular feature, as shown in the circular feature and corresponding centroid on the right. These two sets of points were combined to generate the presence points used in modeling.

Table 1. Occurrence location data by species. Totals indicate the number of occurrence locations for the species derived from WYNDD records. Modeling locations indicates the number of occurrence locations used for model training after filtering for record age and precision. *Species modeled by WYNDD as part of a previous project[8].

Common Name	Scientific Name	Federal Status	Occurrence Locations	Modeling Locations
*Absaroka beardtongue	*Penstemon absarokensis*	Sensitive	160	136
Absaroka biscuitroot	*Lomatium attenuatum*		108	92
Absaroka goldenweed	*Pyrrocoma carthamoides var. subsquarrosa*	Sensitive	108	82
*Beaver Rim phlox	*Phlox pungens*		296	225
*Dubois milkvetch	*Astragalus gilviflorus var. purpureus*		106	43
English sundew	*Drosera anglica*	Sensitive	58	46
*Evert's waferparsnip	*Cymopterus evertii*	Species of Local Concern	58	37
*Fremont bladderpod	*Lesquerella fremontii*	Sensitive	147	94
Hall's fescue	*Festuca hallii*	Sensitive	25	19
Hoary willow	*Salix candida*	Sensitive	108	72
Howard's forget-me-not	*Eritrichium howardii*		34	20
Ice grass	*Phippsia algida*	Species of Local Concern	16	10
Kirkpatrick's ipomopsis	*Ipomopsis spicata var. robruthii*		36	26
Koenigia	*Koenigia islandica*	Species of Local Concern	58	42
Large yellow lady-slipper	*Cypripedium parviflorum var. pubescens*	Sensitive	39	23
Lesser bladderwort	*Utricularia minor*	Sensitive	74	66
Low fleabane	*Erigeron humilis*		20	13
Moschatel	*Adoxa moschatellina*	Species of Local Concern	26	13
Naked-stemmed parrya	*Parrya nudicaulis*	Species of Local Concern	31	17
Narrowleaf goldenweed	*Ericameria discoidea var. linearis*		40	19
North Fork Easter-daisy	*Townsendia condensata var. anomala*	Sensitive	158	140
Oeder's lousewort	*Pedicularis oederi*		46	25

Common Name	Scientific Name	Federal Status	Occurrence Locations	Modeling Locations
Payson's whitlow-grass	*Draba paysonii var. paysonii*		34	24
Rockcress whitlow-grass	*Draba globosa*		46	28
*Rocky Mountain twinpod	*Physaria saximontana var. saximontana*		142	92
Roundleaf orchid	*Amerorchis rotundifolia*	Sensitive	20	10
Russet cottongrass	*Eriophorum chamissonis*	Sensitive	97	72
Sheathed cottongrass	*Eriophorum callitrix*	Species of Local Concern	30	14
*Shoshonea	*Shoshonea pulvinata*	Sensitive	98	51
Snow paintbrush	*Castilleja nivea*		62	50
Sweet-flowered rock jasmine	*Androsace chamaejasme var. carinata*	Species of Local Concern	42	17
Teton wire-lettuce	*Stephanomeria fluminea*		46	36
Trelease's whitlow-grass	*Draba paysonii var. treleasii*		20	14
Twinleaf cinquefoil	*Potentilla subjuga*		16	9
Upward-lobe moonwort	*Botrychium ascendens*	Sensitive	15	14
Weber's saw-wort	*Saussurea weberi*	Species of Local Concern	29	15
White arctic whitlow-grass	*Draba fladnizensis*		29	14
*Wyoming tansymustard	*Descurainia torulosa*	Sensitive	46	30

NEGATIVE DATA COLLECTION AND PROCESSING

True absence data for a given species are seldom available, since most researchers typically do not explicitly record locations where they surveyed for a species but failed to find it. Even when negative results from surveys are recorded, they seldom are databased in a way that makes them readily accessible in numbers sufficient for distribution modeling. Survey routes have routinely been recorded in field notes, but such negative data have not typically been digitized and databased. Given the common lack of robust absence data, methods have been developed to generate "pseudo-absence" or "background" data points[43, 50]. Such methods use background data in order to distinguish between the environmental gradients present in areas that are used by a species versus those in areas that are available to the species. One method for creating a background dataset that is commonly used with the Maxent algorithm for distribution modeling, for example, is to select a large number (e.g., 10,000) of random points from the modeling area to represent the gradients available to a species[51]. However, this

approach assumes that the presence dataset that will be contrasted to the random points is itself a product of random or at least representative sampling.

Botanical initiatives of recent decades, including the pioneering studies of Robert Dorn and the floristic inventories of RM, have aimed for systematic approaches to botanical survey coverage across the state. WYNDD surveys have addressed the need for detailed information on the rarest species. However, these survey initiatives have had discrete scopes and logistical constraints, and nearly all were focused on public lands that sometimes have limited access, so gaps remain in coverage of rare plant presence records.

If sampling bias is not accounted for, a presence-only modeling approach that uses randomized background points may produce a model that predicts sampling effort better than it predicts a species' true distribution[52]. These types of broader-scale sampling biases were addressed in this project by using a target background group approach[50], rather than the default method of selecting random background points or some other means of generating background data. This approach attempts to mirror spatial sampling bias in the presence data for a species by selecting background data – often records for related species – that derive from surveys exhibiting similar spatial biases. Matching the biases in the presence data for a target species with similar biases in the background data helps to factor out systemic sampling bias in modeling, resulting in a model that more accurately reflects a species' distribution. Further, since most of the rare plant presence points used in this modeling project derive from WYNDD botanical surveys, it is reasonable to assume that, had a species been present at any surveyed site, it would have been recorded. Thus, to generate background data for modeling a given species, locations from WYNDD's database for all other species of concern or potential concern were used. Rather than use the gridded points generated for all species, a centroid was generated for the source features representing all plant species of concern or potential concern in Wyoming, to avoid skewing the background dataset to the largest source features. Highly imbalanced sets can result in models that emphasize correct classification of the majority class – absence, in the case of the plant models – over correct classification of the minority class (presences). As such, a down-sampling approach was also used to balance the number of presence and absence points used in modeling (see the "Model Generation" section below, for details).

Environmental Data Organization and Processing

The factors that influence a species' distribution vary across differing spatial scales, from broad-scale gradients like climate to fine-scale parameters such as soil texture[53]. Accordingly, the spatial predictor layers used to build distribution models should represent a similar range of scales in order to produce the most reasonable models[54]. WYNDD has produced an extensive set of environmental predictor layers for distribution modeling, as part of prior modeling projects[4, 5, 7]. These predictors captured important environmental gradients across multiple scales, and related to topography, soils and substrate, hydrology, climate, and vegetation. We explored a subset of 75 of these available predictor layers as potential predictors of distribution for rare plants on the SNF (see Appendix 1 of this report for a full list of potential predictor layers). Although categorical data layers representing mapped units of geology or soils have been used for distribution modeling in some cases, we avoided the use of most categorical datasets because of the propensity for model overfitting when using categorical layers: particularly those with a large number of categories[55]. We evaluated the possibility of using SNF-specific layers that exist, but determined that the benefit of including these more precise data layers would be outweighed in most species' cases by the fact that we would then be unable to use as model training data any occurrence locations outside of the area covered by the layers.

The original set of potential predictors was evaluated for collinearity within Wyoming as part of a prior project[8], and those results were used to reduce the set of potential predictors used to generate distribution models for plants on the SNF to an initial set of candidates. Briefly, we eliminated predictor layers that had a high degree of collinearity with other, more intuitive predictors, across Wyoming. After reducing the potential set of predictors based on collinearity, a set of predictor layers was selected for each species by reviewing all available information on the biology and habitat requirements of the species, erring on the side of initially including predictors that *could be* informative for each species. Random Forest is relatively robust to overfitting[9], and some related evaluations done by WYNDD have shown that beyond an initial variable selection process by a biologist, further variable reduction does not result in substantial differences in models, so no further variable reduction was done as part of this project. We used the Geospatial Modeling Environment (GME)[56] to assign values from each potential predictor layer to the presence and pseudo-absence points. Shapefiles containing these values were then exported as CSV files that could be used in Program R[57]. Models were constructed using all predictor layers identified as potentially informative for each species, based on species biology.

Since using all gridded points for presence source features would amount to pseudoreplication[41], and using all background points would result in class imbalance[58, 59], an iterative approach was taken that used resampling to build many models from many subsets of the presence and absence data. Specifically, for each iteration, a subsample of presence and absence data were selected, and a model was constructed based on those subsamples. Presence point subsamples for the iteration were created by randomly selecting a point from each source feature, to avoid pseudoreplication that would be introduced into the process by taking multiple points from a single source feature. Absence points were randomly "down-sampled"[59] so that there were three times the number of absence points compared to presence points, to mitigate prediction issues due to class imbalance.

After each training set subsample (presence and absence) was drawn, a single Random Forest model was generated by growing 500 trees with the mtry parameter -- the number of predictors to try at each node -- set by default to the square root of the number of total predictors[9]. A total of 100 subsampling iterations were performed, as we found during work producing plant distribution models for the BLM that this number of iterations adequately captured the variability of predictor data values across presence and absence points, leading to stable models[8]. The resulting Random Forest models from each iteration were then combined into a single Random Forest model comprising 50,000 classification trees. A plot of the out-of-bag (OOB) error rate[9] for the combined model resulting from each iteration was generated to further evaluate whether models had stabilized within the number of iterations run. As each of the 100 iterations of a model for a species generated 500 trees, the summary statistics were calculated as the mean of each statistic, based on OOB samples, across all 50,000 trees, using a predicted probability threshold of 50%. Statistics used to evaluate models included the OOB error, True Skill Statistic (TSS)[60], and sensitivity and specificity[61].

Results

Measures of model performance across all species were relatively good (Table 2). OOB error estimates ranged from a low of 0.9% for Fremont bladderpod to 20.4% for Twinleaf cinquefoil, with a median OOB

error of 6.9% across all species. TSS scores, which can range from 0 to 100%, with higher values indicating better models, ranged from a low of 34.7% to a high of 97.0%, respectively, for Twinleaf cinquefoil and Fremont bladderpod. Model performance, as measured by TSS, was weakly correlated with the number of source features available for modeling (R^2 = 0.16), though some species with relatively few points (e.g., Roundleaf orchid, Sheathed cottongrass) had relatively high values for TSS. Roundleaf orchid, for example, had only ten presence locations available for modeling, but had the fifth highest TSS score (93.9%) of all species. This species is extremely narrowly distributed, so it is relatively easy for a modeling algorithm to identify a narrow set of environmental gradients within which the species occurs, resulting in a prediction that tightly fits known distribution. Conversely, Hoary willow, which had a relatively high number of source features available for modeling (72), had relatively poor model performance, especially with respect to sensitivity (80.7%). This could be explained by the fact that the species, though a habitat specialist, is found in most of the major mountain ranges in Wyoming, and more presence data may be needed to adequately represent its distribution.

Table 2. Model performance for final models, by species. Statistics shown here are based on out-of-bag (OOB) samples, using a classification threshold of 50% predicted probability.

Common	OOB Error	TSS	Sensitivity	Specificity	Kappa	Max S+S Threshhold
Absaroka beardtongue	5.7%	84.6%	88.3%	96.3%	84.8%	0.490
Absaroka biscuitroot	5.0%	88.1%	92.2%	95.9%	86.8%	0.638
Absaroka goldenweed	3.8%	89.9%	92.4%	97.5%	89.9%	0.646
Beaver Rim phlox	3.3%	93.4%	96.8%	96.6%	91.3%	0.385
Dubois milkvetch	1.3%	95.5%	95.7%	99.8%	96.6%	0.572
English sundew	9.2%	75.4%	81.5%	93.9%	75.5%	0.556
Evert's waferparsnip	5.8%	83.0%	86.1%	96.9%	84.2%	0.645
Fremont bladderpod	0.9%	97.0%	97.2%	99.8%	97.6%	0.633
Hall's fescue	12.1%	63.3%	69.2%	94.2%	66.3%	0.671
Hoary willow	7.9%	76.7%	80.7%	96.0%	78.6%	0.562
Howard's forget-me-not	9.0%	70.5%	73.8%	96.7%	74.6%	0.553
Ice grass	7.4%	82.8%	88.9%	93.9%	80.8%	0.771
Kirkpatrick's ipomopsis	12.7%	67.0%	75.9%	91.1%	66.5%	0.636
Koenigia	2.4%	94.6%	96.8%	97.9%	93.6%	0.490
Large yellow lady-slipper	7.7%	77.6%	81.8%	95.8%	79.0%	0.643
Lesser bladderwort	7.0%	80.4%	84.7%	95.7%	81.1%	0.661
Low fleabane	13.4%	64.3%	73.3%	91.0%	64.2%	0.596
Moschatel	14.9%	85.1%	58.5%	94.0%	57.0%	0.545
Naked-stemmed parrya	7.2%	81.8%	87.1%	94.7%	81.0%	0.472
Narrowleaf goldenweed	6.4%	79.4%	81.8%	97.6%	82.3%	0.783
North Fork Easter-daisy	5.1%	86.8%	90.3%	96.5%	86.6%	0.608
Oeder's lousewort	2.3%	95.3%	97.4%	97.8%	94.0%	0.543
Payson's whitlow-grass	10.8%	73.3%	81.5%	91.8%	71.8%	0.520
Rockcress whitlow-grass	9.6%	77.4%	85.2%	92.1%	75.2%	0.732
Rocky Mountain twinpod	4.2%	88.5%	91.1%	97.4%	88.9%	0.397

Common	OOB Error	TSS	Sensitivity	Specificity	Kappa	Max S+S Threshhold
Roundleaf orchid	2.7%	93.9%	96.3%	97.6%	92.9%	0.561
Russet cottongrass	6.8%	80.6%	84.4%	96.2%	81.7%	0.638
Sheathed cottongrass	4.8%	91.7%	97.2%	94.5%	87.7%	0.813
Shoshonea	4.0%	88.9%	91.4%	97.5%	89.2%	0.463
Snow paintbrush	8.3%	79.5%	85.8%	93.7%	78.2%	0.703
Sweet-flowered rock jasmine	5.6%	86.2%	90.4%	95.8%	85.3%	0.657
Teton wire-lettuce	3.3%	91.5%	93.8%	97.7%	91.3%	0.322
Trelease's whitlow-grass	12.1%	69.0%	77.6%	91.4%	68.2%	0.627
Twinleaf cinquefoil	20.4%	34.7%	42.9%	91.8%	38.9%	0.713
Upward-lobe moonwort	9.8%	69.8%	74.3%	95.5%	72.8%	0.620
Weber's saw-wort	5.5%	86.4%	90.5%	95.9%	85.6%	0.360
White Arctic Whitlow-grass	11.7%	68.5%	76.3%	92.3%	68.7%	0.747
Wyoming tansymustard	7.4%	81.1%	86.5%	94.6%	80.3%	0.543

Climate predictors were included in all species' models, and had an mean importance score (MDA) of 0.019. Bioclim4 (Temperature Seasonality) and bioclim10 (Mean Temperature of Warmest Quarter) were particularly important to species models, with mean importance scores of 0.033 and 0.027, respectively. Among other predictors, Soil cation-exchange capacity (cec_surf), slope, and growing degree days (growdd) appeared in models for fewer species (7, 1, and 8, respectively), but had had relatively high importance values for those species' models (0.034, 0.033, and 0.032, respectively). Variables based on transformations of slope aspect values were of relatively low importance, with all four of these predictors having mean MDA scores of 0.001 or below.

Table 3. Predictor layers included in modeling, with the number of final models that included the predictor and the average importance (Mean Decrease Accuracy) of that predictor across models including that predictor.

Name	Predictor	Number of Models	Average Importance
A[1] (Transformed Aspect -- Southeast/Northwest Gradient)	aprime135	11	0.001
A[1] (Transformed Aspect -- North/South Gradient)	aprime180	11	0.001
A[1] (Transformed Aspect -- Southwest/Northeast Gradient)	aprime45	11	0.001
A[1] (Transformed Aspect -- West/East Gradient)	aprime90	9	0.000
Available water capacity, surface soil layer	awc_surf	11	0.006
Bare ground cover	bare	23	0.004
Mean Temperature of Warmest Quarter	bioclim10	38	0.027
Annual Precipitation	bioclim12	38	0.021
Precipitation of Wettest Month	bioclim13	38	0.016
Precipitation Seasonality (Coefficient of Variation)	bioclim15	38	0.012
Precipitation of Driest Quarter	bioclim17	38	0.019

Name	Predictor	Number of Models	Average Importance
Precipitation of Warmest Quarter	bioclim18	38	0.016
Mean Diurnal Range (Mean of monthly (max temp - min temp))	bioclim2	38	0.015
Isothermality (BIO2/BIO7) (* 100)	bioclim3	38	0.018
Temperature Seasonality (standard deviation *100)	bioclim4	38	0.033
Min Temperature of Coldest Month	bioclim6	38	0.013
Biome	biome	38	0.007
Soil cation-exchange capacity, surface soil layer	cec_surf	7	0.034
Landscape Contagion Index	contag	4	0.003
Compound Topographic Index	cti	12	0.004
Distance to Cliffs	d2cliffs40	14	0.018
Distance to Rock Outcrop	d2outcrop	14	0.004
Distance to Permanent Flowing Water	d2pfw	3	0.011
Distance to Wetland Habitat	d2wethab	10	0.026
Depth to water table	dep2watr	10	0.003
Soil Electrical Conductivity top 200 cm	ec_surf	6	0.007
Elevation	elev	19	0.023
Flooding Frequency Class	flood_freq	1	0.002
Forest Canopy Cover	forestcc	3	0.019
Frost Days	frostdays	8	0.023
Calcium rating of bedrock geology formation	geol_calc	18	0.009
Growing Degree Days	growdd	8	0.032
Herbaceous Cover Index	herb	1	0.023
Heat Load Index	hli	9	0.003
Soil organic matter, surface soil layer	orgmatsurf	7	0.010
Percent clay, surface soil layer	pclaysurf	3	0.005
Cottonwood Index	pode	1	0.004
Percent sand, surface soil layer	psandsurf	21	0.017
Percent silt, surface soil layer	psiltsurf	23	0.015
Degree Slope	slope	1	0.033
Soil pH, surface soil layer	soilphsurf	6	0.016
Topographic Position Index, 11-cell focal window	tpi_11	26	0.007
Topographic Position Index, 3-cell focal window	tpi_3	26	0.005
Topographic Position Index, 31-cell focal window	tpi_31	24	0.009
Vector Ruggedness Measure, 11-cell focal window	vrm11	14	0.010
Vector Ruggedness Measure, 3-cell focal window	vrm3	14	0.004
Vector Ruggedness Measure, 31-cell focal window	vrm31	14	0.016

DISCUSSION

USAGE AND LIMITATIONS OF DISTRIBUTION MODELING

The quality and utility of predictive distribution models is necessarily limited by the quality and quantity of presence/absence data, the availability of important predictor layers, and the uncertainties associated with the biogeographic processes that influence realized distribution. Presence data are subject to error in recording, reporting, and data entry, and misidentification may be an issue with species for which identification is difficult. Available presence data may also be the result of uneven sampling effort, which can result in patterns of occurrence locations that are more an artifact of sampling than a manifestation of habitat preferences by the species. Likewise, predictor data layers may contain errors, or may fail to adequately represent the important environmental gradients that drive the complex biological processes that shape distribution.

For example, the presence data for Twinleaf cinquefoil draws almost completely on collection records. It is a species that has been regarded as common in Colorado[62] so has not warranted consideration as a sensitive species in the USFS Rocky Mountain Region or focus of survey work in SNF. However, the work of a Flora of North America expert has provided evidence that it may represent an undescribed species restricted to SNF[63]. The coarseness of this species' presence data may account for the exceptionally low TSS value of its model.

The species addressed in this project represent many interesting cases, and even with precise presence data and ideal environmental layers, it is problematic to model species such as Moschatel. It is distributed in both the Black Hills and in the Greater Yellowstone plus one location in the Bighorn Mountains, all with similar microhabitat conditions but fundamentally different climate envelopes and landscape differences. In other words, this species is present in the Black Hills at elevations that are mostly below 6000 feet and are surrounded by the Great Plains, as well as in the Greater Yellowstone landscapes mostly above 6000 feet in the middle of the Rocky Mountains.

For many rare plant species in Wyoming, distribution appears to be a function of both available habitat and processes associated with geographic isolation. Thus, models like those prepared in this project represent "potential distribution." Despite the presence of error and uncertainty, such models remain useful hypotheses about a species' geographic distribution, as long as users understand the inherent limitations of each model.

MODEL INTERPRETATION AND USAGE

Potential distribution models are tools, and in the case of SNF, valuable additions to existing tools represented by the environmental data layers already developed for the Forest. Although the output values from distribution models are commonly thought of as a logistic probability of a species' presence, the actual interpretation is typically more nuanced. Without substantial and representative absence data, it is impossible to determine the species' extent across large landscapes[64], since it is unknown whether empty spots on the "dot maps" of species observations are truly unoccupied. Thus, there is no direct way to estimate the true probability of a species' presence at any given location. Instead, output values from models such as the ones developed for this project should be viewed as relative indices of suitability for a

species. Output from two different models cannot be directly compared (i.e., a value of 0.5 in a model for one species may not mean the same thing as a value of 0.5 in another species' model, and the true probability of a species occurring in such a location may not be 50% in either case). Higher output values should generally correspond with a higher probability of presence, and vice-versa, so models can be used to identify the areas that have the highest potential for species' presence.

Distribution models such as those produced in this project can help identify the species of interest that are *potentially* present in a proposed project area. They might also identify areas of potentially high concentrations of target species, at a coarse scale. Planners can use such maps to help them determine areas that may be more suitable for development or management activities with less likelihood of adverse impacts to biodiversity or to a particular species[42].

Distribution models can also be used to guide field surveys. By selecting the areas predicted by a model to be most suitable, researchers can hone in on the most likely locations to find a particular species to make the most of limited field project budgets. Moreover, by evaluating model output in the context of known presence points, researchers can focus on areas a model deems suitable but that currently have no known records for the species, potentially expanding its known distribution. However, models should not be used in place of site-level, clearance surveys for TES species, as the predictor layers used to create distribution models are generally too coarse to make an accurate prediction at this scale. For project planning at a site-level, models can provide only an indication of whether the species is predicted to be "in the neighborhood," in which case field surveys may be warranted.

Final model products from this project were delivered as continuous, predicted suitability values, as a binary (predicted present/predicted absence) representation, and as simplified output showing four ordinal categories of suitability. Any use of a distribution model may require expressing the model differently by applying different thresholds or symbology in mapping the model output. A biologist interested in locating a particular species, for example, would most benefit by limiting their sampling to only the areas predicted to be most highly suitable for their target species (i.e., focusing on only the top-most category – "Highest predicted probability of presence"). Conversely, a manager tasked with evaluating the potential impact of development for a species or group of species may want to err on the side of caution, by considering even areas of lower predicted likelihood of presence to be potentially occupied and warranting field surveys (i.e., initially ruling out only the "Predicted Absent" category). Any additional information that may be available on a local scale, such as detailed soils or geology data for the SNF, can be used in conjunction with the models as another indicator of suitability.

Caution must also be exercised when evaluating the partial plots generated for each species' model. These plots are show the relative likelihood of presence as a function of each variable, holding all other variables constant. While indirect predictor layers such as elevation might contribute substantially to the accuracy of a model, it would not be correct in most cases to state, for example, that elevation has a specific effect on distribution. Rather, elevation most likely influences temperature, precipitation, vegetation, soils, or other gradients that more directly limit a species' distribution. Biological understanding is thus important in interpreting partial plots – particularly those for more indirect predictors[65].

Occurrence Data Limitations

Over half of the species modeled as part of this project were known from fewer than 30 distinct locations. Thirty locations has sometimes been suggested as a practical lower limit for producing predictive distribution models[42], though other have produced useful models with far fewer presence locations[42-49]. Models for species with fewer than 30 presence locations available for model building, in particular, should be considered an initial hypothesis about potential distribution for the species. Some of the species represented by these models may have distributions that extend beyond the currently known distributions; in these cases, substantially better models may result if additional, independent observations are made in expanded portions of the species' distributions and added to the modeling sets for these species.

Predictor Data Limitations

For most plants, soil characteristics are extremely important in limiting distribution, and this has been recognized for native plants in Wyoming[66]. Unfortunately, detailed digital soils data layers (SSURGO) are not currently available as a statewide coverage for Wyoming, and likely will not become available for a number of years (J. Bauchert, pers. comm.). Although portions of the state, including most of the SNF, have coverage of the detailed soils layers, making use of the data in a subset of the study area introduces large areas where no predictions can be made, as they lack the necessary predictor values. Statewide soils layers provide some information that can help in model building, but completion of the more detailed, SSURGO soil data layer would allow much more precise predictions to be made. Users of the models can add SSURGO data, or other, more detailed predictor layers available for the SNF or other locales in the state, in combination with model output, as additional context for model usage.

Species that were found in several mountain ranges around the state tend to have lower model performance scores, on average, than those that occur in relatively isolated and contiguous areas. This reflects the relative ease with which a modeling algorithm can identify narrow ranges of environmental gradients within which the species occur, for very narrowly distributed species. Hoary willow is also a habitat specialist but even more widely distributed across an even broader geographic area, across a range of climate envelopes and elevations. For these species, the results suggest the limitations of modeling unless sectors of its distribution were modeled separately.

Suggestions for Future Work

Additional presence locations for the modeled species have the potential to improve distribution models. New locations that are identified far from previous locations, or in novel habitats, have the greatest potential for impacting the results of any subsequent modeling. Substantial and representative absence data for the modeled species likewise could improve distribution models, by enabling true presence/absence modeling techniques that could produce superior models to those produced with presence/background or presence/pseudo-absence methods.

Absence data can be collected directly, when a species is surveyed for but not found, or it can be created retroactively based on prior survey work that found other species, but not the target species. While creating pseudo-absence data from locations where other species were recorded seems reasonable in this case, explicitly building negative datasets for each species could greatly improve models, particularly

at fine scales. WYNDD has recently created a new database that allows for the structured storage and retrieval of negative data. Consolidating the negative records that currently exist as assorted GIS files and printed records into the new database would help lay the groundwork for a future modeling effort using true presence/absence methods.

As with presence point data, collection or generation of newer and better predictor datasets should continue to be a priority for modeling work. This includes refinement of existing data layers, and development of new data layers based on remotely sensed data that are made available on a regular basis as satellite imagery becomes more ubiquitous. Detailed soil data, in particular, is likely to be important in refining models for rare plants in Wyoming, as it eventually becomes available throughout the state.

ACKNOWLEDGEMENTS

REFERENCES

[1] U.S. Department of Agriculture, Forest Service. (2016) Threatened, endangered and sensitive plants and animals. Supplement No: 2600-2015-1. In: FSM 2600 – Wildlife, fish and sensitive plant habitat management chapter 2670. Effective Date: August 29, 2015. (http://www.fs.usda.gov/detail/r2/landmanagement/) [Accessed 26 February 2016].

[2] U.S. Department of Agriculture, Forest Service. (2015) U.S. Department of Agriculture, Forest Service [USDA FS]. 2015a. Land Management Plan – 2015 Revision. Shoshone National Forest. Cody, WY. http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprd3842886.pdf. [Accessed 26 February 2016]. Cody, WY.

[3] Fertig, W., and Thurston, R. (2003) Modeling the potential distribution of BLM Sensitive and USFWS Threatened and Endangered plant species in Wyoming. Report prepared for the Bureau of Land Management Wyoming State Office by the Wyoming Natural Diversity Database., Laramie, WY.

[4] Beauvais, G. P., Andersen, M. D., and Keinath, D. A. (2012) Range, distribution, and habitat of terrestrial vertebrates in the 5-state Northwest ReGAP region. Report prepared for the USDI Geological Survey - Gap Analysis Program (Moscow, Idaho), University of Wyoming, Laramie, Wyoming.

[5] Keinath, D., Andersen, M., and Beauvais, G. (2010) Range and modeled distribution of Wyoming's species of greatest conservation need, Report prepared by the Wyoming Natural Diversity Database, Laramie Wyoming for the Wyoming Game and Fish Department, Cheyenne, Wyoming and the US Geological Survey, Fort Collins, Colorado.

[6] Heidel, B., Handley, J., and Andersen, M. (2011) Distribution and habitat requirements of Desert Yellowhead (*Yermo xanthocephalus*), Fremont County, Wyoming.

[7] Keinath, D. A., Griscom, H. R., and Andersen, M. D. (2014) Habitat and distribution of the Wyoming pocket gopher (*Thomomys clusius*), *Journal of Mammalogy 95*, 803-813.

[8] Andersen, M., Heidel, B., and Beauvais, G. (2016) Predictive Distribution Modeling of BLM Threatened, Endangered, and Sensitive Plant Species in Wyoming, Wyoming Natural Diversity Database, University of Wyoming, Laramie, WY.

[9] Breiman, L. (2001) Random forests, *Machine learning 45*, 5-32.

[10] Heidel, B. (2012) Wyoming plant species of concern, Wyoming Natural Diversity Database, University of Wyoming, Laramie, WY.

[11] Kirkpatrick, R. S. (1987) A Flora of Southeastern Absarokas, Wyoming., Masters thesis, Department of Botany, University of Wyoming, Laramie, WY.

[12] Snow, N. (1994) The vascular flora of southeastern Yellowstone National Park and the headwaters region of the Yellowstone River, Wyoming, *The Wasmann Journal of Biology 50*, 52-95.

[13] Massatti, R. T. (2007) A floristic inventory of the East Slope of the Wind River Mountain Range and vicinity, Wyoming, Masters Thesis. Department of Botany, University of Wyoming, Laramie, WY.

[14] Elliott, E. R. (2014) A floristic inventory of the northern Absaroka, Beartooth, and Gallatin Ranges, Wyoming and Montana, U.S.A., Masters Thesis. Department of Botany, University of Wyoming, Laramie, WY.

[15] Evert, E. F. (2010) *Vascular Plants of the Greater Yellowstone Area: Annotated Catalog and Atlas*, Erwin Evert, Park Ridge, IL.

[16] Scott, R. W. (1997) *The Alpine Flora of the Rocky Mountains; Volume 1. The Middle Rockies*, University of Utah Press, Salt Lake City, UT.

[17] Fertig, W., and Jones, G. (1992) Plant communities and rare plants of the Swamp Lake Botanical Area, Clarks Fork Ranger District, Shoshone National Forest, Unpublished report prepared for the Shoshone National Forest by the Wyoming Natural Diversity Database, Laramie, WY.

[18] Fertig, W., and Bynum, M. (1994) Biological report on the Proposed Bald Ridge Research Natural Area, Unpublished report prepared for the Shoshone National Forest by the Wyoming Natural Diversity Database, Laramie, WY.

[19] Fertig, W., and Bynum, M. (1994) Biological Report on the Proposed Picketts Knob Research Natural Area., Prepared for the Shoshone National Forest by the Wyoming Natural Diversity Database, Laramie, Wyoming.

[20] Jones, G. P., and Fertig, W. (1999) Ecological evaluation of the potential Arrow Mountain Research Natural Area within the Shoshone National Forest, Fremont County, Wyoming, Unpublished report prepared for the Shoshone National Forest, USDA Forest Service by the Wyoming Natural Diversity Database, Laramie, WY.

[21] Jones, G. P., and Fertig, W. (1999) Ecological evaluation of the potential Beartooth Butte Research Natural Area within the Shoshone National Forest, Park County, Wyoming., Prepared for the Shoshone National Forest, USDA Forest Service by the Wyoming Natural Diversity Database, Laramie, WY.

[22] Jones, G. P., and Fertig, W. (1999) Ecological evaluation of the potential Grizzly Creek Research Natural Area within the Shoshone National Forest, Park County, Wyoming, Unpublished report prepared for the Shoshone National Forest, USDA Forest Service by the Wyoming Natural Diversity Database, Laramie, WY.

[23] Jones, G. P., and Fertig, W. (1999) Ecological evaluation of the potential Lake Creek Research Natural Area within the Shoshone National Forest, Park County, Wyoming, Unpublished report prepared for the Shoshone National Forest, USDA Forest Service by the Wyoming Natural Diversity Database, Laramie, WY.

[24] Jones, G. P., and Fertig, W. (1999) Ecological evaluation of the potential Pat O'Hara Mountain Research Natural Area within the Shoshone National Forest, Park County, Wyoming,

Unpublished report prepared for the Shoshone National Forest, USDA Forest Service by the Wyoming Natural Diversity Database, Laramie, WY.

[25] Jones, G. P., and Fertig, W. (1999) Ecological evaluation of the potential Sheep Mesa Research Natural Area within the Shoshone National Forest, Park County, Wyoming, Unpublished report prepared for the Shoshone National Forest, USDA Forest Service by the Wyoming Natural Diversity Database, Laramie, WY.

[26] Fertig, W. (1998) The status of rare plants on Shoshone National Forest: 1995-97 survey results, Unpublished report prepared for the Shoshone National Forest by the Wyoming Natural Diversity Database, Laramie, WY.

[27] Dorn, R. D. (1989) A report on the status of *Lomatium attenuatum*, a Candidate Threatened species, Prepared for the U.S. Fish and Wildlife Service by Mountain West Environmental Services, Cheyenne, WY.

[28] Dorn, R. D. (1989) A report on the status of *Penstemon absarokensis*, a Candidate Threatened species, Unpublished report prepared for the U.S. Fish and Wildlife Service by Mountain West Environmental Services, Cheyenne, WY.

[29] Dorn, R. D. (1989) Report on the status of *Descurainia torulosa*, a Candidate Threatened species, Unpublished report prepared for the U.S. Fish and Wildlife Service by Mountain West Environmental Services, Cheyenne, WY.

[30] Dorn, R. D. (1989) Report on the status of *Shoshonea pulvinata*, a Candidate Threatened species, Unpublished report prepared for the U.S. Fish and Wildlife Service by Mountain West Environmental Services, Cheyenne, WY.

[31] Fertig, W. (1995) Status report on *Parrya nudicaulis* in Bridger-Teton National Forest, Unpublished report prepared for the Bridger-Teton National Forest by the Wyoming Natural Diversity Database, Laramie, WY.

[32] Fertig, W., Welp, L., and Markow, S. (1999) Status report on Evert's waferparsnip (*Cymopterus evertii*) in northwestern Wyoming, Unpublished report prepared for the Bureau of Land Management Wyoming State Office by the Wyoming Natural Diversity Database, Laramie, WY.

[33] Heidel, B. (2009) Status of *Lesquerella macrocarpa* (large-fruited bladderpod) and *Phlox pungens* (Beaver Rim phlox) in the Upper Green River Basin, Unpublished report prepared for the Bureau of Land Management Pinedale Field Office by the Wyoming Natural Diversity Database, Laramie, WY.

[34] Heidel, B. (2011) Status of *Shoshonea pulvinata* (Shoshonea), Park County, Wyoming, Unpublished report prepared for the Bureau of Land Management, Cody Field Office by the Wyoming Natural Diversity Database, Laramie, WY.

[35] Heidel, B. (2013) Sensitive plant survey in the Dickinson Park area, Shoshone National Forest, Unpublished report prepared for the Shoshone National Forest by the Wyoming Natural Diversity Database, Laramie, WY.

[36] Markow, S. (2001) Report on a survey for *Stephanomeria fluminea* in Grand Teton National Park, Unpublished document prepared for the Greater Yellowstone Inventory and Monitoring Project by the Wyoming Natural Diversity Database, Laramie, WY.

[37] Markow, S. (2004) Survey for *Stephanomeria fluminea* on the Bridger-Teton National Forest, Unpublished report prepared for Bridger-Teton National Forest and the Region 4 Forest Service Office by the Wyoming Natural Diversity Database, Laramie, WY.

[38] Marriott, H. J. (1991) Survey for *Androsace chamaejasme var. carinata* and *Saussurea weberi*, northwest Wind River Range, Wyoming, Unpublished report prepared for the Bridger-Teton National Forest by the Wyoming Natural Diversity Database, Laramie, WY.

[39] Marriott, H. J. (1992) Status report for *Descurainia torulosa* (Wyoming tansymustard), Sweetwater County, WY, Unpublished report prepared for the Bureau of Land Management, Rock Springs District by the Wyoming Natural Diversity Database, Laramie, WY.

[40] Taylor, A., Taylor, K., Heidel, B., and Handley, J. (2005) North Fork sensitive plant surveys on Shoshone National Forest, Unpublished report prepared for the Shoshone National Forest by the Wyoming Natural Diversity Database, Laramie, WY.

[41] Hurlbert, S. H. (1984) Pseudoreplication and the design of ecological field experiments, *Ecological monographs 54*, 187-211.

[42] Franklin, J., and Miller, J. A. (2009) *Mapping species distributions: spatial inference and prediction*, Vol. 338, Cambridge University Press Cambridge.

[43] Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J. R., Lehman, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., and Moritz, C. (2006) Novel methods improve prediction of species' distributions from occurrence data, *Ecography 29*, 129-151.

[44] Hernandez, P. A., Graham, C. H., Master, L. L., and and Albert, D. L. (2006) The effects of sample size and species characteristics on performance of different species distribution modeling methods, *Ecography 29*, 773-785.

[45] Stockwell, D. R., and Peterson, A. T. (2002) Effects of sample size on accuracy of species distribution models, *Ecological modelling 148*, 1-13.

[46] Kadmon, R., Farber, O., and Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models, *Ecological Applications 13*, 853-867.

[47] Loiselle, B. A., Jørgensen, P. M., Consiglio, T., Jiménez, I., Blake, J. G., Lohmann, L. G., and Montiel, O. M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes?, *Journal of Biogeography 35*, 105-116.

[48] Kremen, C., Cameron, A., Moilanen, A., Phillips, S., Thomas, C., Beentje, H., Dransfield, J., Fisher, B., Glaw, F., and Good, T. (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools, *Science 320*, 222-226.

[49] Wisz, M. S., Hijmans, R., Li, J., Peterson, A. T., Graham, C., and Guisan, A. (2008) Effects of sample size on the performance of species distribution models, *Diversity and Distributions 14*, 763-773.

[50] Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., and Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data, *Ecological Applications 19*, 181-197.

[51] Phillips, S. J., and Dudik, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation, *Ecography 31*, 161-175.

[52] Newmark, W. D. (1996) Insularization of Tanzanian parks and the local extinction of large mammals, *Conservation Biology 10*, 1549-1556.

[53] Johnson, D. H. (1980) The comparison of usage and availability measurements for evaluating resource preference, *Ecology 61*, 65-71.

[54] Meyer, C. B., and Thuiller, W. (2006) Accuracy of resource selection functions across spatial scales, *Diversity and Distributions 12*, 288-297.

[55] Hastie, T., Tibshirani, R., and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second ed., Springer, New York, New York, United States of America.

[56] Beyer, H. L. (2012) Geospatial Modelling Environment (Version 0.7.2.1), Hawthorne L. Beyer, URL: http://www.spatialecology.com/gme.

[57] R Core Team. (2013) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Available at: http://www.R-project.org, Vienna, Austria.

[58] Chawla, N. V., Japkowicz, N., and Kotcz, A. (2004) Editorial: special issue on learning from imbalanced data sets, *ACM Sigkdd Explorations Newsletter 6*, 1-6.

[59] Chen, C., Liaw, A., and Breiman, L. (2004) Using random forest to learn imbalanced data, *University of California, Berkeley*.

[60] Allouche, O., Tsoar, A., and and Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS), *Journal of Applied Ecology 43*, 1223-1232.

[61] Fielding, A. H., and Bell, J. F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models, *Environmental Conservation 24*, 38-49.

[62] Ackerfield, J. (2015) *Flora of Colorado.*, Botanical Research Institute of Texas Press, Fort Worth, Texas.

[63] Ertter, B., Elven, R., Reveal, J. L., and Murray, D. F. (2014) *Potentilla.*, Vol. Vol. 9, pp. 121-218., Oxford University Press, New York and Oxford.

[64] Hastie, T., and Fithian, W. (2013) Inference from presence-only data; the ongoing controversy, *Ecography 36*, 864-867.

[65] Austin, M., Belbin, L., Meyers, J., Doherty, M., and Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory, *Ecological Modelling 199*, 197-216.

[66] Fertig, W. (2011) Strategies for plant conservation in Wyoming: Distributional modeling, gap analysis, and identifying species at risk, In *Department of Botany*, University of Wyoming, Laramie, Wyoming.