GREEN DIAMOND Resource Company

Modeling Habitat Suitability of *Monotropa uniflora* using Maxent (a Maximum Entropy Approach)

David W. Lamphear, Gabe Cashman, Macy Budesilich, and Alexandria Shedlock

Introduction

Monotropa uniflora, known commonly as Indian-pipe or ghost-pipe, is an herbaceous perennial with circumboreal distribution. Its global range includes temperate and boreal forests of North America, Asia, European Russia and northern South America. In North America, the species is known to occur in 11 Canadian provinces and 44 states within the continental U.S. (Figure 1). The California Natural Diversity Database global ranking, which reflects the overall status of a species throughout its global range, is G5 or "demonstrably secure"; however, the state ranking is S2 or "imperiled" due to the restricted range within the state border. The California Native Plant Society currently classifies the species as a California Rare Plant Rank 2B.2; this status acknowledges that the *M. uniflora* has a limited geographic range in the State but is (globally) common elsewhere. This species is neither State nor Federally Listed as Endangered, Threatened or Rare.

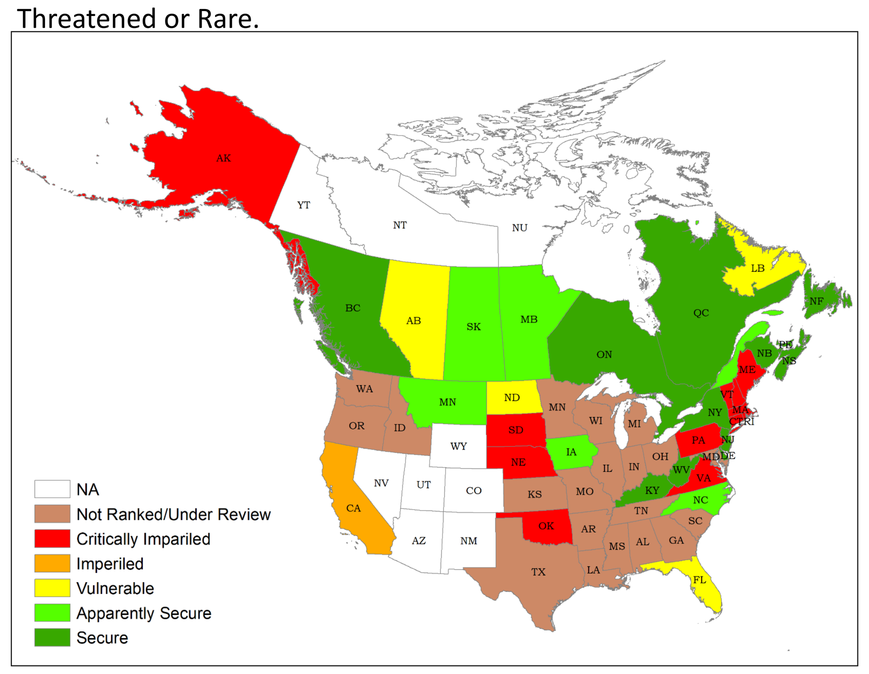


Figure 1. Distribution and State Ranking for *M. uniflora (NatureServe 2024).*

In California, *M. uniflora* occurs in Del Norte and northern Humboldt counties in mixed coniferous stands where Douglas-fir (*Pseudotsuga menziesii*) is the dominant overstory species. Similar to other members of the Monotropoideae, *M. uniflora* is mycoheterotrophic, which is defined as plants with the ability to obtain carbon resources from associated mycorrhizal fungi which are linked to autotrophic plants. Unlike green (photosynthetic/autotrophic) plants, *M. uniflora* is distinguished by its bright white pigmentation (achlorophyllus), solitary terminal flower and a propensity to occur in the dark understory of northern temperate forests. Like most mycoheterotrophic plants, *M. uniflora* associates with a small range of fungal hosts: all of them members of Russulaceae.

Figure 2. Flowers typically emerge from the ground in a nodding position then become upright when fertilized.

Figure 3. As the species name suggests, the stems bear only a single flower which is 10–15 mm long with 5 (3-8) free sepals which are more or less identical to the 5 (3-8) free petals.

Figure 4. Reproductive stems often reach heights of 10–30 (cm) with small scale-like leaves 5–10 (mm) long.

Figure 5. Reproduction is facilitated by small, dust-like seeds that are wind-dispersed.

Monotropa uniflora on Green Diamond Property

Green Diamond Resource Company (GDRCo) is committed to managing special status plant species across their ownership. To this end, all proposed timber harvest areas undergo extensive botanical surveys focusing on special status plants which are either 'State and Federally Listed Endangered, Threatened, and Rare Plants of California' or included in the 'California Rare Plant Ranking' (CRPR) system maintained by the California Native Plant Society (CNPS).

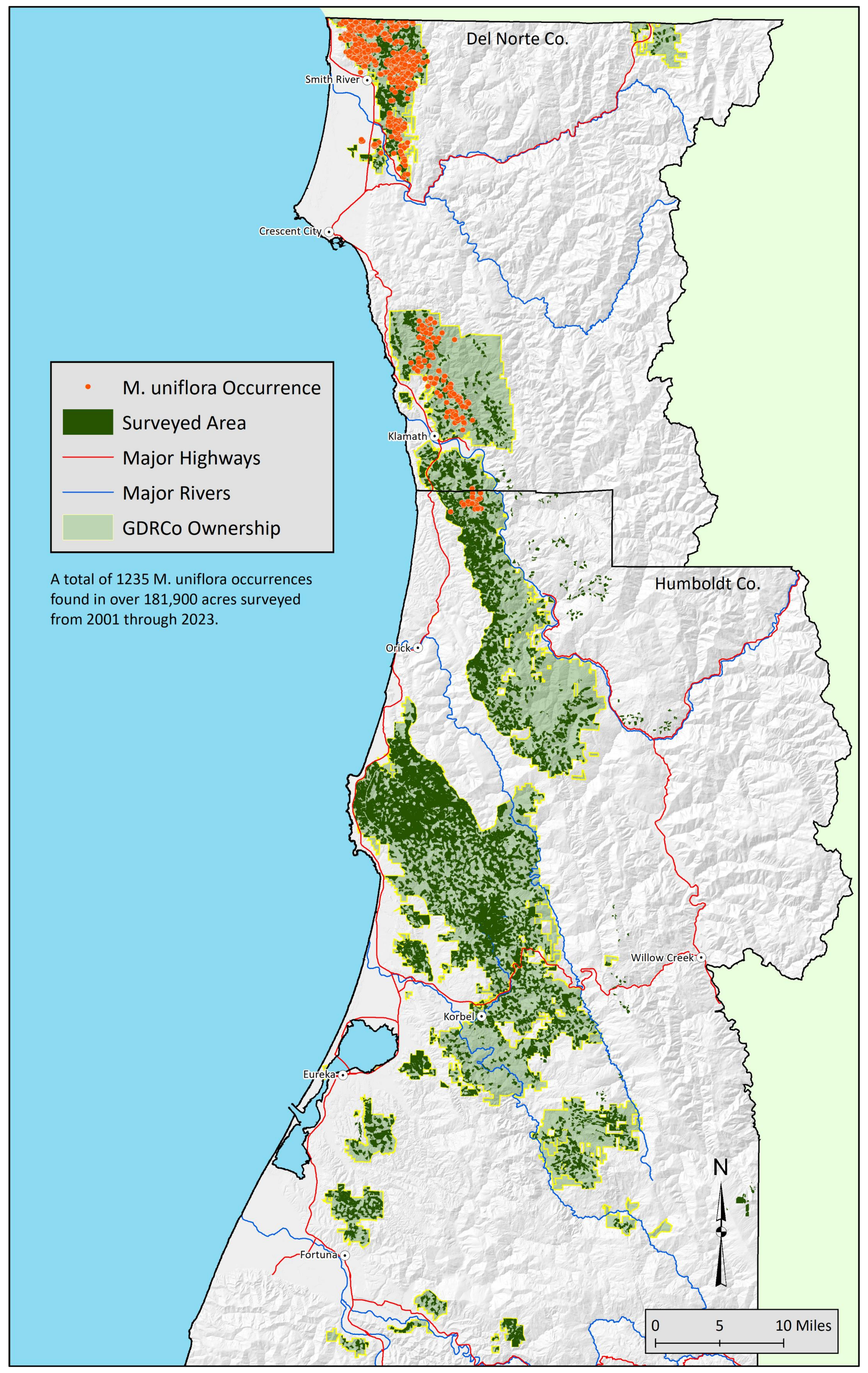


Figure 6. *M. uniflora* occurrences and surveyed areas on Green Diamond Property

During pre-harvest botanical surveys (2001 - 2023), *M. uniflora* was found at 1,235 locations on GDRCo property within the 238,297 ha Study Area (GD Property plus ½ mi buffer). As minimal records (39) exist outside GDRCo property the modeling effort was constrained to company property with known survey effort in Humboldt and Del Norte Counties (Figure 6).

M. uniflora is commonly found in patches of one to several hundred plants distributed throughout a timber harvest unit. While these groupings are mapped as unique locations, their close proximity to one-another poses a modeling risk known as spatial autocorrelation and therefore, may not be independent samples. To reduce this potential bias, a minimum separation distance of 100 meters was used to subsample the occurrence data set.

In addition to entering multiple botanical species into the GDRCo GIS, all areas where botanical surveys are conducted are mapped in detail. Project Surveys Areas (PSA) become important within the Maxent modeling process.

Species Distribution Modeling With Maxent

Maxent is a maximum entropy-based modeling approach that utilizes occurrence-only data with raster variables to predict habitat suitability for a target species. Utilizing established Use vs Availability techniques, Maxent is a thoroughly proven, well recognized, and robust species distribution modeling method. Maxent is particularly suited for small sample sizes and data sets with occurrence-only records.

Maximum Entropy

In ordinary language, the principle of maximum entropy can be said to express a claim of epistemic modesty, or of maximum ignorance. The selected distribution is the one that makes the least claim to being informed beyond the stated prior data. By choosing to use the distribution with the maximum entropy allowed by our information, the argument goes; we are choosing the most uninformative distribution possible. To choose a distribution with lower entropy would be to assume information we do not possess. Thus, the maximum entropy distribution is the only reasonable distribution. (Unnamed source)

Modeling of habitat suitability for *Monotropa uniflora* was constrained to GDRCo property within Del Norte and Humboldt Counties. While minimal occurrence records are located outside GDRCo ownership, it was deemed important to assess suitability in an area with known and quantified survey effort.

Within the area of interest GDRCo has developed over 80, 5-meter resolution, raster- based covariate data sets. These include biographic and geographic (elevation, slope, aspect, coast distance), climatological (precipitation and temperature) and physical (solar radiation, soils, etc.) covariates to mention a few. Twenty-six predictor variables were chosen for this effort.

Model Selection

Model selection was accomplished in three phases using a best subsets approach (Table 2). Maxent software options included 10 bootstrap replicates, random seed, regularization multiplier of two, 5000 iterations, scaled logistic output, and model form combinations including Linear, Quadratic, Product, and Hinge. All models in Phases 1-3 were trained using a random subset of 80 percent of presence locations and tested using the remaining 20 percent. The background (available) dataset was constructed from a Balanced Acceptance Sample (BAS) of 115,000 locations within the Study Area.

We evaluated pairwise variable correlations for all predictors in program R using Pearson's coefficient, Cramer's V, and box plots. In Phases 2-3, variables with Pearson's coefficient, and Cramer's V values above 0.60 were excluded from entering the model at the same time. The HSM models we estimated are known to be robust to the effects of covariate collinearity during estimation and variable selection (Elith et al., 2011, Feng et al., 2019). A list of the variables used is found in Table 1.

We assessed the predictive strength of individual covariates using a combination of marginal plots, percent contribution, permutation importance, and regularized training and testing gain.

Table 1. Predictive variables assessed for inclusion in final model

Variable Theme	Variable Name	Variable Description
Biographic/Geographic	Elev	Elevation in meters above mean sea level (NAVD88)
	Slope_p	Slope in percent (first derivative of elevation (Elev)
	Aspect_4d	Catagorical aspect represented by North 315-45 (1), South 135-225 (2), East 225-315 (3), and West 45-135 (4)
	Latitude	Latitude in degrees north
	Sp2000	Relative elevation-based position between the ridge top and the valley bottom directly downslope
	TPI_2000	Topographic Position Index run with a circular radius of 2,000 meters
	TPI_1000	Topographic Position Index run with a circular radius of 1,000 meters
	TPI_300	Topographic Position Index run with a circular radius of 300 meters
	TPI_150	Topographic Position Index run with a circular radius of 150 meters
	Coast_Dist	Distance inland from the Pacific Ocean in meters
	EucDist_PerStrms	Euclidian distance from the nearest perennial (Class I and Class II) stream
	AreaSol	Combination of diffuse and direct incoming solar radiation in watthours per square meter (WH/m2)
	AreaSolDir	Direct incoming solar radiation in watthours per square meter (WH/m2)
	AreaSolDif	Diffuse incoming solar radiation in watthours per square meter (WH/m2)
	AreaSolDirDur	Duration of direct incoming solar radiation in hours
	ZCI	Modeled inland extent of coastal fog and the Pacific Ocean moderating effect on higher inland air temperatures.
Climate	PPT_30yr_Normal	PRISM 30-yr normal (1991-2020) annual precipitation (mm)
	TMAX_30yr_Normal	PRISM 30-yr normal (1991-2020) annual average maximum temperature (Celsius)
	TMEAN_30yr_Normal	PRISM 30-yr normal (1991-2020) annual average mean temperature (Celsius)
	TMIN_30yr_Normal	PRISM 30-yr normal (1991-2020) annual average minimum temperature (Celsius)
Soils	pH_Surf_Hor_W_Avg	pH as a weighted average of the surface horizon. A numerical expression of the relative acidity or alkalinity of a soil sample. (SSM)
	Depth_Surf_Hor_W_Avg	Depth as a weighted average of the surface horizon. Distance from the soil surface to the base of the top soil horizon.
	Organic_Matter_Surf_Hor_W_Avg	Organic matter as a weighted average of the surface horizon. The amount by weight of decomposed plant and animal residue expressed as a weight percentage of the less than 2 mm soil material.
	Hydraulic_Cond_Surf_Hor_W_Avg	Hydraulic Conductivity as a weighted average of the surface horizon. The amount of water that would move vertically through a unit area of saturated soil in unit time under unit hydraulic gradient.
	Bulk_Density_Surf_Hor_W_Avg	Bulk density as a weighted average of the surface horizon. The oven dry weight of the less than 2 mm soil material per unit volume of soil exclusive of the desication cracks, measured on a coated clod.
	Avail_H2O_Cont_Surf_Hor_W_Avg	Available water content as a weighted average of the surface horizon. The amount of water that an increment of soil depth, inclusive of fragments, can store that is available to plants. AWC is expressed as a volume fraction, and is commonly estimated as the difference between the water contents at 1/10 or 1/3 bar (field capacity) and 15 bars (permanent wilting point) tension and adjusted for salinity, and fragments.

Covariates with high predictive strength exhibited a biologically reasonable marginal response and high relative influence (typically values ≥ 20), as measured by percent contribution and permutation importance. The top 2 to 4 covariates from each group depending on differences in predictive strength advanced to phase 2.

During Phase 1, we estimated independent models in Maxent for each of the 3 covariate groups to rank predictive ability within groups. The Biographic/Geographic group was included twice, once with and without the Latitude covariate.

During phase 2, we combined all top covariates from Phase 1 but prohibited covariates with pairwise Pearson correlation greater than 0.60 from appearing in the same model. This resulted in four HSM models and the top 3-4 covariates from these four models moved on to Phase 3.

During phase 3, poorly performing covariates from Phase 3 were dropped and the models were estimated. We designated these four candidate models based on each group A, B, C, and D and selected the model with highest predictive strength as measured by boxplots of AIC calculated over 10 bootstrap iterations.

Table 2. Model Phases showing model variable groupings, Percent Contribution, and Permutation Importance.

Phase 1

Biographic Geographic			Biographic Geographic (sans			Climate			Soils		
Variable	PC	PI	Variable	PC	PI	Variable	PC	PI	Variable	PC	PI
Latitude	87.4	70.2	Coast_Dist	58.0	49.4	* PPT_30yr_Norm	71.0	73.9	Depth	28.5	12.5
Coast_Dist	4.6	22.3	* ZCI	21.6	13.5	TMEAN_30yr_Norm	13.6	15.4	pH	23.6	35.4
Elev	4.5	5.8	* Elev	6.3	8.5	* TMAX_30yr_Norm	7.7	5.7	Hydraulic_Cond	15.6	4.2
sp2000	0.8	0.1	AreaSolDif	3.2	15.8	TMIN_30yr_Norm	5.2	3.7	Avail_H2O_Cont	15.6	14.4
Aspect_4d	0.6	0.0	EucDist_PerStrms	2.1	0.9	ZCI	2.5	1.3	Bulk_Density	13.6	15.5
AreaSolDirDur	0.5	0.2	sp2000	1.9	0.5				Organic_Matter	3.1	18.0
TPI_2000	0.4	0.6	AreaSolDirDur	1.9	4.1						
AreaSolDif	0.4	0.1	Aspect_4d	1.3	0.1						
Slope_p	0.2	0.3	TPI_2000	1.2	2.4						
EucDist_PerStrms	0.2	0.1	Slope_p	1.0	3.0						
AreaSolDir	0.2	0.1	TPI_1000	0.7	0.4						
ZCI	0.1	0.0	AreaSolDir	0.5	0.8						
TPI_1000	0.1	0.1	TPI_150	0.1	0.2						
TPI_150	0.0	0.0	AreaSol	0.1	0.4						
TPI_300	0.0	0.1	TPI_300	0.1	0.0						
AreaSol	0.0	0.0									

Phase 2

A			B			C			D		
Variable	PC	PI	Variable	PC	PI	Variable	PC	PI	Variable	PC	PI
Latitude	79.9	40.8	Latitude	82.7	70.1	PPT_30yr_Norm	48.1	46.0	PPT_30yr_Norm	50.9	65.1
PPT_30yr_Norm	13.9	36.3	PPT_30yr_Norm	13.2	24.6	Coast_Dist	34.2	41.2	Elev	24.6	19.0
Coast_Dist	3.8	20.9	Elev	2.3	2.5	TMEAN_30yr_Norm	7.6	8.2	TMEAN_30yr_Norm	12.9	11.5
AreaSolDif	1.1	1.0	AreaSolDif	0.9	0.9	ZCI	6.8	2.7	AreaSolDif	4.7	2.2
TMEAN_30yr_Norm	0.6	0.7	Depth	0.4	0.3	AreaSolDif	2.2	1.0	ZCI	3.6	0.5
Depth	0.5	0.3	TMEAN_30yr_Norm	0.3	1.3	Depth	1.1	0.7	Depth	2.8	1.0
ZCI	0.1	0.0	ZCI	0.2	0.2	pH	0.1	0.2	pH	0.5	0.8
pH	0.1	0.0	pH	0.0	0.0						

Phase 3

A (AUC = 0.971)			B (AUC = 0.971)			C (AUC = 0.965)			D (AUC = 0.965)		
Variable	PC	PI	Variable	PC	PI	Variable	PC	PI	Variable	PC	PI
Latitude	81.1	36.8	Latitude	84.3	75.5	PPT_30yr_Norm	49.4	41.7	PPT_30yr_Norm	54.7	66.4
PPT_30yr_Norm	14.4	35.0	PPT_30yr_Norm	12.8	22.5	Coast_Dist	35.5	45.4	Elev	26.4	20.7
Coast_Dist	4.4	28.2	Elev	2.9	1.9	TMEAN_30yr_Norm	7.8	8.8	TMEAN_30yr_Norm	13.0	9.6
						ZCI	7.2	4.1	AreaSolDif	5.9	3.3

Model Evaluation

Initially, we used receiver operating characteristic curve (ROC) and area under curve (AUC) output from Maxent as an index to evaluate overall model prediction ability. AUC ranges from 0 to 1, with a value of 0.5 indicating predictions no different than random chance and with AUC =1 indicating perfect discrimination. Our final HSM model had an AUC value of 0.971 (Table 2, Phase 3). Furthermore, AUC varied little over the ten bootstrap replicates evaluated by the Maxent software (Figure 7). This value indicates a model with high predictive ability and low probability of false negatives.

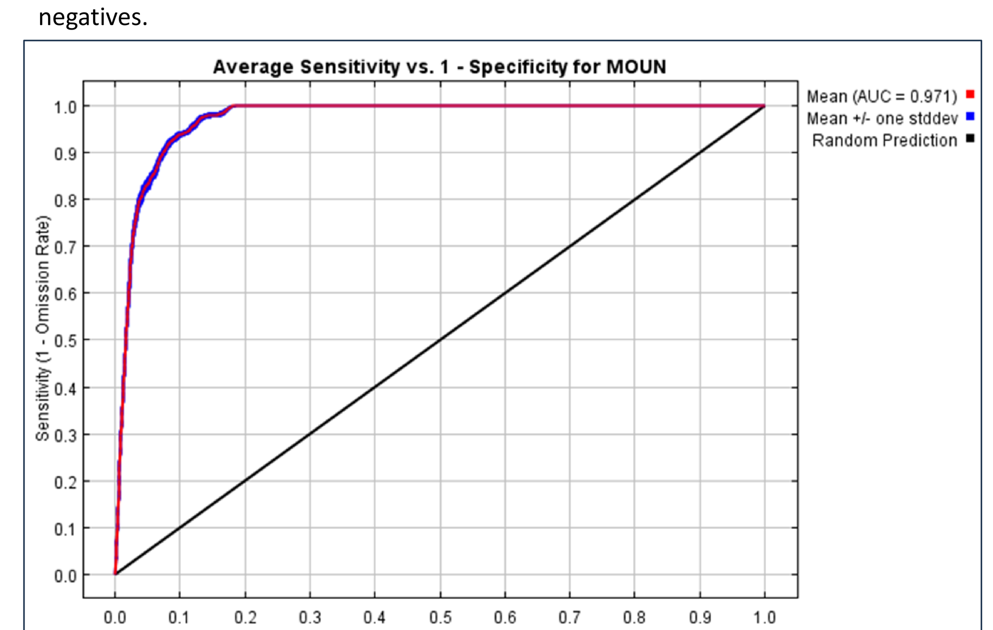


Figure 7. Receiver operating characteristic (ROC) curve averaged over 10 replicate runs. This graph compares model sensitivity (true positive rate) against specificity (true negative rate). Our top model (A), predicting the environmental niche of *M. uniflora* yielded an AUC value of 0.971, signifying Omission/Commission rates are very low.

The top Phase 3 model was selected by evaluating averaged AIC values and boxplots. The results for each model, listed in Table 3 below and ranked by AIC value. AIC evaluates model fit and applies a penalty based on the number of parameters in the model. Model A exhibited both the lowest AIC value and least AIC dispersion between replicates followed by a non-competitive model C, ΔAIC of 824, Figure 8.

Table 3. Ranking of top four models based on Akaike Information Criteria, the best model A consisted of Latitude, Precipitation, and Distance from Coast

Model	Average of nParams	Average of logLikelihood	Average of AIC	Delta AIC
A	70.0	-19,937	40,015	
C	65.0	-20,355	40,839	824
D	77.2	-20,848	41,851	1,836
B	61.5	-20,904	41,932	1,917

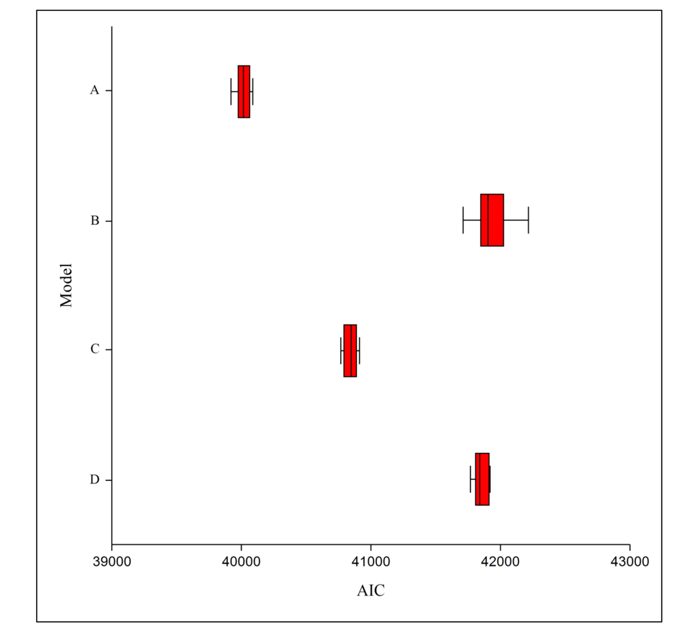
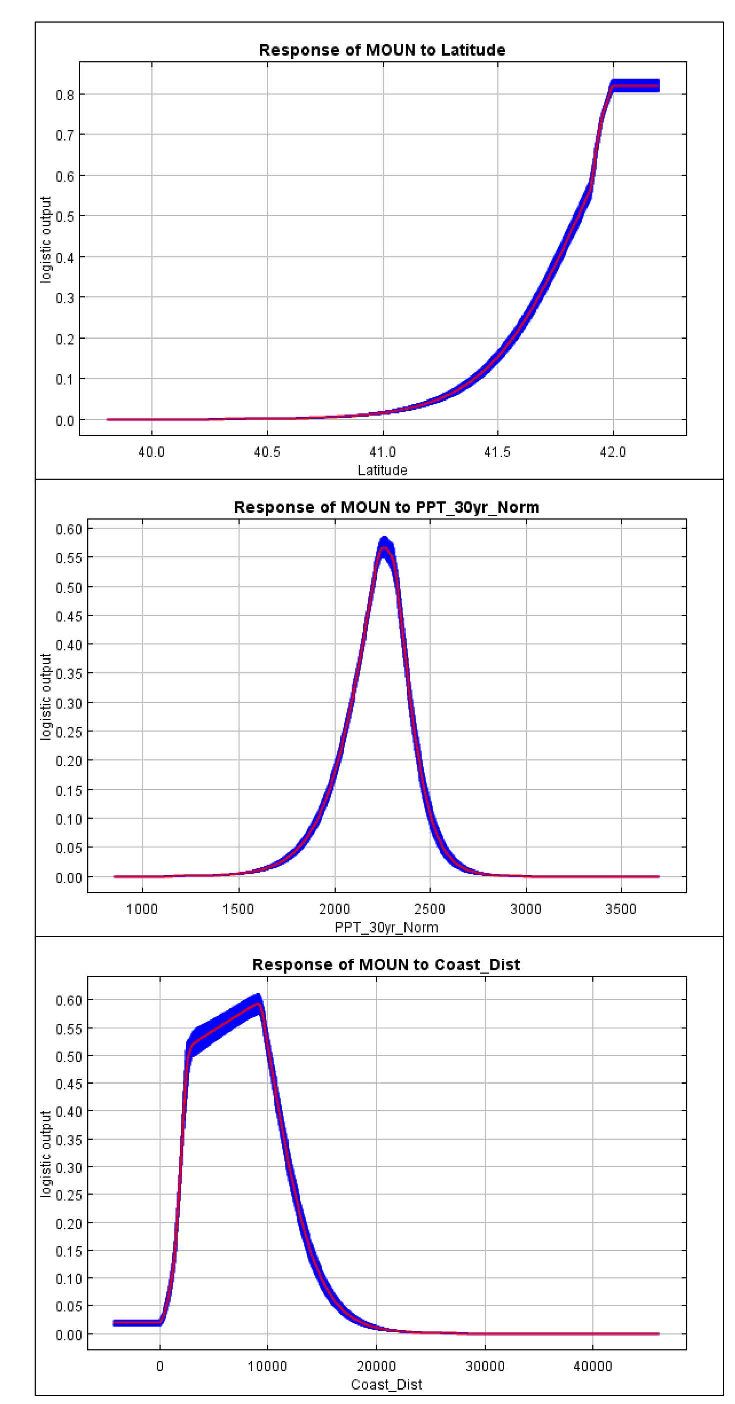


Figure 8. Box and whisker plot of Akaike Information Criteria (AIC) values for top 4 candidate models. Median and dispersion for model A are the lowest suggesting both parsimony and consistency between replicates. Box boundaries are at the 25th and 75th percentiles, the whiskers extend from the 10th to the 25th and the 75th to the 90th percentiles.

Marginal response curves illustrate the effect of each predictor variable within the model (Figures 9 – 11). In general, *M. uniflora* prefers the northern latitudes of the property, generally higher yearly precipitation between 2163 and 2373 mm, and distance to the coast of less than 11 km.



Figures 9-11. *M. uniflora marginal* response to latitude (degreed north), precipitation (mm), and distance from coast (km).

The final model predictions, shown in Figure 12, represent the relative probability of occurrence of *M. uniflora* distributed across the study area. Generally high suitability is localized in the most northern Smith River tract and the upper- and mid-Klamath regions. A lower probability area protruding from the northeast to the upper-middle portion of the Smith River area is most likely due to significantly higher rainfall and is evident in the gap in distribution of *M. uniflora* in Figure 6.

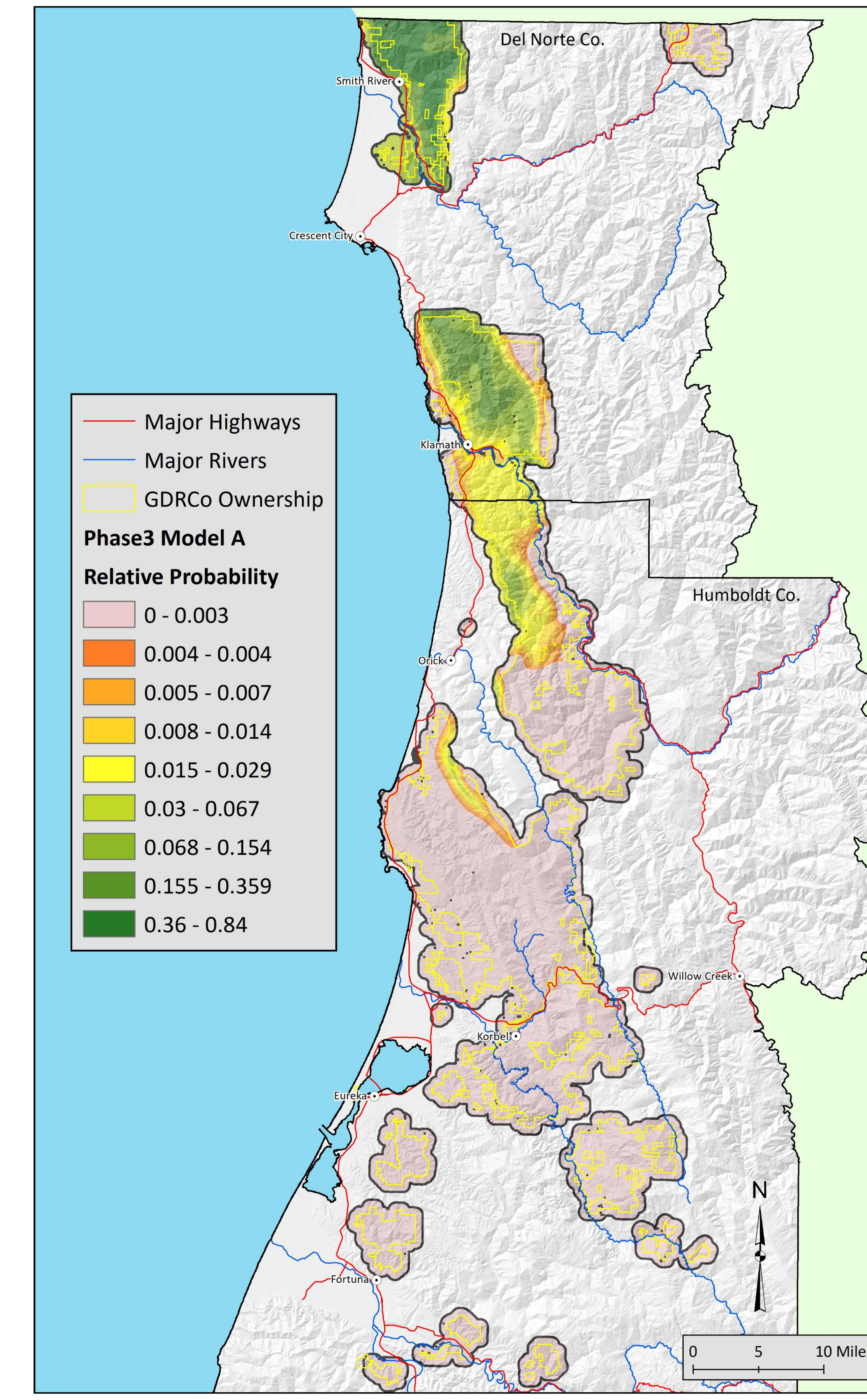


Figure 12. Relative probability of *M. uniflora* occurrence across the study area in northwestern California.

Assessing the distribution of predicted relative probability of occurrence between the background values (115,000 locations) and the known presence (1,235 locations) illustrates a model's ability to separate the habitat niche from unoccupied or very low-quality habitat (Figure 13).

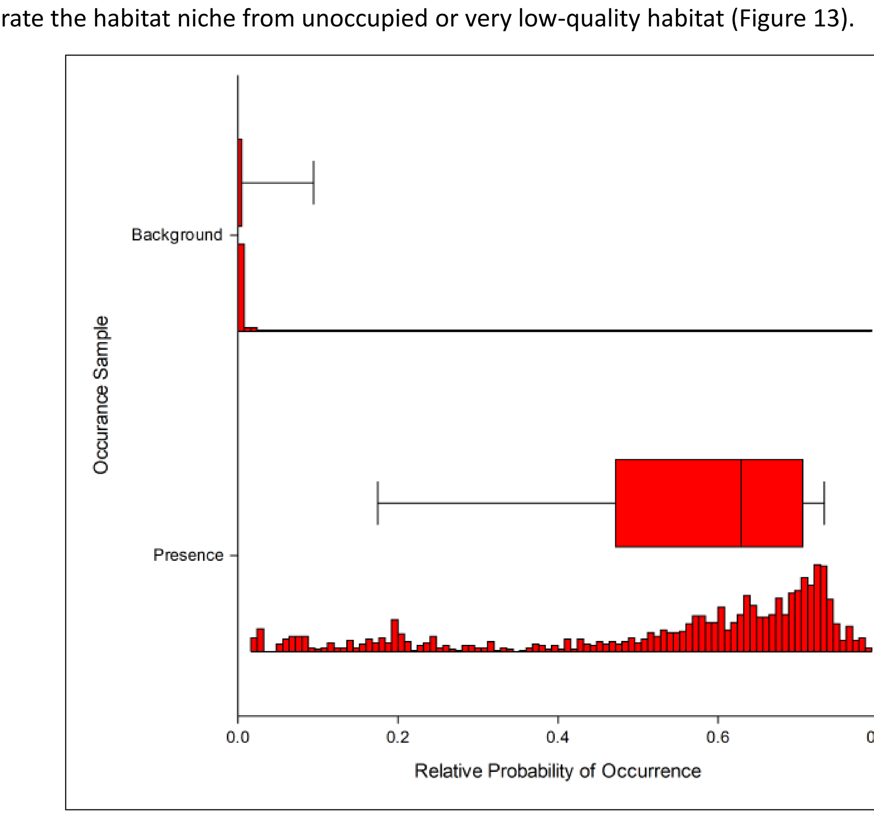


Figure 13. Comparative distribution plots of relative probability of *M. uniflora* occurrence for Presence and Background samples across the study area.

Figure 14. Typical forest habitat associated with *M. uniflora*

Acknowledgements

We would like to thank E. Goldsworthy, B. Hayashi and C. Sanville, prior botanists at Green Diamond, for their time in collecting, developing, and analyzing data in support of this project. We are also grateful for all the work put in by all the botanical technicians involved in the field surveys and mapping efforts so critical to determining a species distribution.