

Species Distribution Modeling of the Channel Islands Tree Poppy (*Dendromecon harfordii*)

*Multivariate generalized linear model of the distribution of Channel Islands Tree Poppy (*Dendromecon harfordii*) using elevation, slope, and aspect as predicting variables of habitat*

Hilary and Skyler



Abstract

The Channel Island Tree Poppy (*Dendromecon harfordii*) endemic to the Channel Islands off southern California is listed as a vulnerable species by the State of California and is threatened by feral herbivores and climate change. The goal of this project was to identify parameters that drive *D. harfordii* distribution in the Channel Islands and model habitat suitability to predict areas where this species may occur in order to inform conservation efforts onto which areas of the island to focus efforts on and to predict the effects of climate change on this species. We developed a multivariate generalized linear model (GLM) using the GIS toolbox MGET that uses elevation, slope, and aspect to predict the occurrence and distribution of *D. harfordii*. Species distribution data was used from GBIF and elevation, slope, and aspect data were derived from a 10-meter DEM. Our results showed that elevation was the only significant predictor of *D. harfordii* presence (p-value < 0.5). Our model was relatively weak in predicting the occurrence of *D. harfordii* (accuracy=0.729, error rate=0.271). Future work should be done to determine other variables that may be better predictors or may be combined to make a better model of species distribution.

Introduction

The Channel Island Tree Poppy (*Dendromecon harfordii*, formerly *Dendromecon rigida* subsp. *harfordii*) is a small perennial shrub or tree that is endemic to the Channel Islands off southern California, currently inhabiting Santa Catalina, Santa Cruz, and Santa Rosa Islands (1). The Channel Island Tree Poppy is part of the poppy family (Papaveraceae) and has bluish-green oval-shaped leaves and bright yellow flowers with four petals (2). It grows from 2-6 meters tall and blooms between March and November (3,4). The Channel Island Tree Poppy is found mostly in the chaparral or coastal sage scrub habitats in course, well-drained soil between 15-420 meters in elevation (4).

The Channel Island Tree Poppy has a state rank (S-rank) of S3, meaning it is "vulnerable in the state due to a restricted range, relatively few populations (often 80 or fewer), recent and widespread declines, or other factors making it vulnerable to extirpation"(4). Current threats to the species are feral herbivores (i.e. sheep) and climate change, which may affect the habitat suitability and conditions that the Channel Island Tree Poppy needs to exist (4). Because of its range is restricted due to its island geography, there is less ability for the species to adapt to climate change by altering its range, making this species more vulnerable to the impacts of climate change.

The goal of this project is to identify parameters that drive *D. harfordii* distribution in order to model habitat suitability and predict areas where this species may occur. To do so, we developed a multivariate model including elevation, slope, and aspect as parameters that influence *D. harfordii* distribution to predict the occurrence of *D. harfordii*.

Conceptual model: presence of D. harfordii ~ slope + elevation + aspect

Materials and Methods

Data for the spatial distribution (presence) of *D. harfordii* was acquired from the open access biodiversity database GBIF (5). Data was downloaded from GBIF using *rgbif* code in R. Inaccurate data was removed in ArcGIS using two methods: clip from the draw tool and manual removal in attribute table. Data was removed included points not on the Channel Islands (i.e. points in the middle of the Atlantic ocean, points in the water of the Southern California Bight, and points on mainland USA). Because the distribution of *D. harfordii* is on small islands, instead of using a buffer to create absence points as was recommended, we used a random point generator to create absence points of *D. harfordii* (first having to transform our raster of the islands to a polygon so that points were just created on land).

Data for the predictor variables (elevation, slope, and aspect) were acquired a 10-meter resolution DEM raster layer from the National Elevation Dataset (6). Elevation of the Channel Islands was directly given by the DEM and we then used ArcGIS tools to create both slope and aspect rasters from the DEM layer.

We then used the Marine Geospatial Ecology Tools (MGET) toolbox, an open source geoprocessing toolbox that can be used in ArcGIS, to: perform preliminary statistical analysis of our variables; fit a generalized linear model (GLM) using our elevation, slope, and aspect layers; evaluate the strength of our GLM model and the strength of our variables as predictors; and create a predictive map for the presence of *D. harfordii* based on the probability of occurrence Conceptual model: presence of *D. harfordii* ~ slope + elevation + aspectce.

Results

A Cleveland plot (Fig. 1) shows that there are no trends to the slope, elevation, and aspect data, revealing that no transformations on those variables are needed. A scatterplot matrix of those variables (Fig. 2) also shows no trends, revealing that no collinearity exist between the variables.

A generalized linear model (GLM), using elevation, slope, and aspect as predictors of presence and absence of *D. harfordii* is as follows (Fig. 3):

$$Presence = - 0.0079471 * slope + 0.0016613 * elevation - 0.005368 * aspect$$

In this equation, 1 is presence and 0 is absence of *D. harfordii*. Slope is measured in percent, elevation in meters and aspect in degrees (0 and 360 = N).

Only elevation significantly predicts the occurrence of *D. harfordii* (p-value < 0.05), whereas slope and aspect are not indicated as significant predictors of *D. harfordii* (p-values > 0.05) (Fig. 3). The Variance Inflation Factors (VIF) are all below 3 in value, meaning that there is no collinearity between variables occurring (Fig. 3). The cutoff value for prediction is 0.259 and the area under the ROC curve is 0.484 (Fig. 4, 5). Both of those values are rather low, meaning that the probability of false positives is not very low and that our model is therefore relatively weak. The rates of true and false positive and negatives are shown further in Fig. 4 and the Cohen's kappa (K) value of 0.110 also reveals our model is weak (Fig. 4). The accuracy of the model is 0.729 and the error rate is 0.271 (Fig. 4). Overall, these results show that our model is relatively weak.

The predictive map generated in ArcGIS (Fig. 6) again indicates some weakness in our model. Areas predicted to have high likelihood of species occurrence do not always have a strong correlation with actual observations. When looking on Santa Cruz Island, the area of highest probability is clearly the high ridgeline running through the center of the island. However, there are no observations of *D. harfordii* in this area. In addition, low lying areas close to the shoreline are predicted to have low probability of *D. harfordii* occurrences but have many physical observations. When looking at the Catalina Island *D. harfordii* population, the map appears to have better correlation between prediction and observation with a large population of *D. harfordii* on the ridges in the center of the island. Overall, the results of our predictive map show enough correlation between slope, elevation and aspect to be a useful tool. However, the amount of diversion between prediction and observation leaves the question of if there is another predictor variable that plays a significant role in *D. harfordii* distribution.

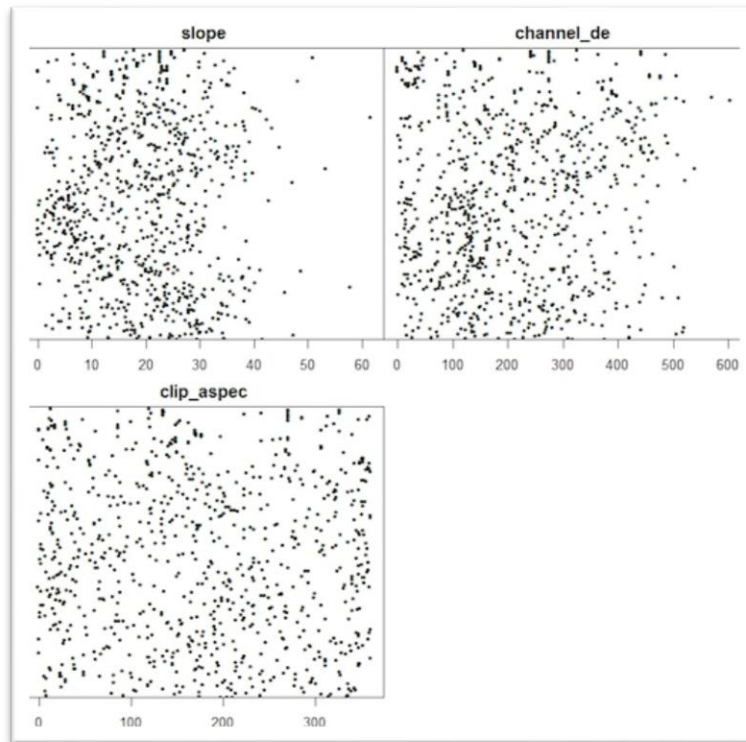


Figure 1. Cleveland Plot of slope, elevation, and aspect parameters in the Channel Islands based on a 10-meter DEM layer (USGS).

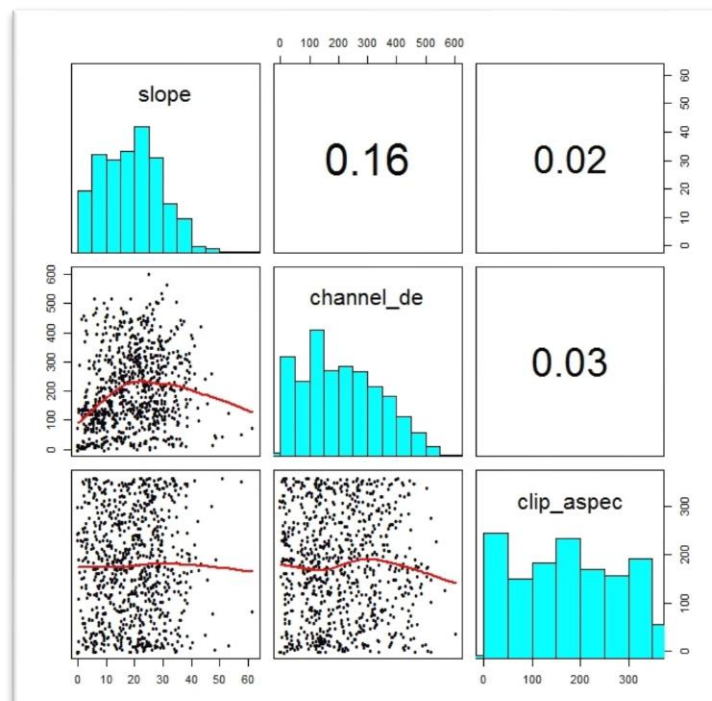


Figure 2. Scatterplot Matrix of slope, elevation, and aspect parameters in the Channel Islands based on a 10-meter DEM layer (USGS).

```

MODEL SUMMARY:
=====

Call:
glm(formula = Presence ~ slope + channel_de + clip_aspec, family = binomial(link="logit"), data = na.omit(d))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9153  -0.7210  -0.6667  -0.5982   1.9520

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3874838   0.2907988  -4.771 1.83e-06 ***
slope        -0.0079471   0.0100404  -0.792   0.4286
channel_de    0.0016613   0.0007583   2.191   0.0285 *
clip_aspec   -0.0005368   0.0009465  -0.567   0.5706
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 614.71  on 584  degrees of freedom
Residual deviance: 609.51  on 581  degrees of freedom
AIC: 617.51

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit
Response: Presence

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                584      614.71
slope    1    0.1992      583      614.51 0.65535
channel_de 1    4.6773      582      609.83 0.03056 *
clip_aspec 1    0.3220      581      609.51 0.57043
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Deviance explained = 0.8%

Variance Inflation Factors:

              VIF
channel_de 1.035678
slope      1.033951
clip_aspec 1.003077

```

Figure 3. Summary output of the GLM using elevation, slope, and aspect as predictors of the presence and absence of *D. harfordii* in the Channel Islands.


```

=====
Model Performance Summary:

Area under the ROC curve (auc)           = 0.484
Mean cross-entropy (mxe)                 = 0.509
Precision-recall break-even point (prbe) = 0.254
Root-mean square error (rmse)            = 0.404

Cutoff selected by maximizing the Youden index = 0.259

Confusion matrix for that cutoff:

          Actual 1  Actual 0  Total
Predicted 1      15      35      50
Predicted 0      44     198     242
Total            59     233     292

Model performance statistics for that cutoff:

Accuracy (acc)           = 0.729
Error rate (err)         = 0.271
Rate of positive predictions (rpp) = 0.171
Rate of negative predictions (rnp) = 0.829

True positive rate (tpr, or sensitivity) = 0.254
False positive rate (fpr, or fallout)    = 0.150
True negative rate (tnr, or specificity)  = 0.850
False negative rate (fnr, or miss)        = 0.746

Positive prediction value (ppv, or precision) = 0.300
Negative prediction value (npv)               = 0.818
Prediction-conditioned fallout (pcfall)       = 0.700
Prediction-conditioned miss (pcmiss)          = 0.182

Matthews correlation coefficient (mcc)        = 0.111
Odds ratio (odds)                           = 1.929
SAR                                           = 0.539

Cohen's kappa (K)                           = 0.110

```

Figure 4. Model performance summary of the GLM using elevation, slope, and aspect as predictors of the presence and absence of *D. harfordii* in the Channel Islands.

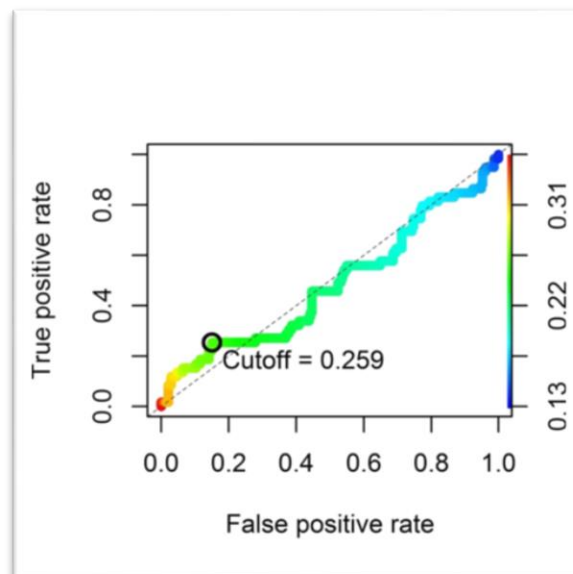


Figure 5. ROC curve of the GLM using elevation, slope, and aspect as predictors of the presence and absence of *D. harfordii* in the Channel Islands, showing a cutoff value of 0.259 and the area under ROC as 0.484.

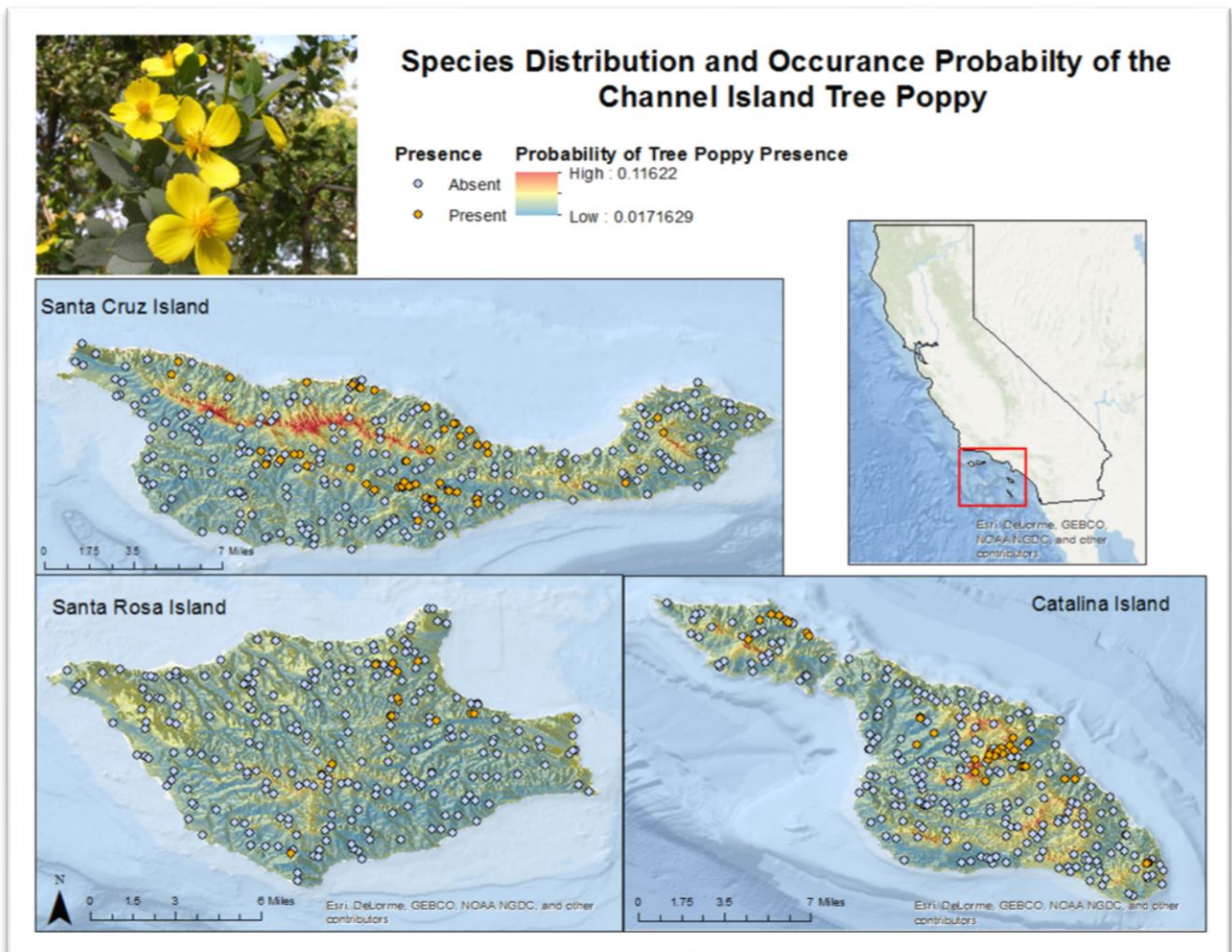


Figure 6. A predictive species observation map illustrates the calculated probability of *D. harfordii* presence based on terrain slope, elevation and aspect. Presence and absence data points allow us to evaluate the strength of the model through the correlation of predicted probability and actual observations. Slope, aspect and elevation derived from data provided by the USGS.

Discussion

The primary threats to the Channel Island Tree poppy are feral herbivores and climate change (4). The results of this project can be used to predict the effects climate change will have on this species as the suitability of habitat changes with a changing climate. The results may also be used to inform conservation efforts (i.e. restoration or feral pig or sheep removal) on which areas of the island to focus efforts on in order to protect this vulnerable species. Such areas can be identified by utilizing the probability map.

However, the results from our model and the discrepancy between the generated probability of presence and the actual observations of presence lead us to conclude that our model is relatively weak. Because of this, future work needs to be done to test and determine other variables that may be better predictors of *D. harfordii* habitat. Potential variables may include soil type, precipitation, habitat type, proximity to ocean (salt water content), proximity to stream or ravine and proximity to other individuals. By exploring these additional predictor variables, we would be able to produce a model with stronger predictions of *D. harfordii* presence. In addition, by exploring more climatic variables, we would be better able to predict how climate change could potentially affect *D. harfordii* distribution moving into the future. With these additional variables, the model can then be better used to inform land management and conservation efforts to protect the viable population range moving into the future.

Literature Cited

1. Calflora. 2014. "Taxon Report 2685." Accessed Oct 14 2014: http://www.calflora.org/cgi-bin/species_query.cgi?where-taxon=Dendromecon+rigida+ssp.+harfordii.
2. Golden West College/. 2014. "Plant of the Month: October 2010." Accessed Oct 14, 2014: <http://www.goldenwestcollege.edu/garden/plantofmonth/1010.html>.
3. Wikipedia. 2014. "*Dendromecon harfordii*." Accessed Oct 14, 2014: http://en.wikipedia.org/wiki/Dendromecon_harfordii.
4. CNPS, Rare Plant Program. 2014. "*Dendromecon harfordii* var. *harfordii*." Inventory of Rare and Endangered Plants. Accessed Oct 14, 2014: <http://www.rareplants.cnps.org/detail/561.html>.
5. GBIF. 2014. "*Dendromecon harfordii* Kellogg." Accessed Nov 7, 2014: <http://www.gbif.org/species/2888425>.
6. National Elevation Dataset, USGS. Last Modified April, 15, 2014.