Habitat Characterization of Gaviota Tarplant (*Dienandra increscens ssp. villosa*)

Benjamin D. Best, Jake Marcon, Kathleen Dayton

2021-03-02 15:00

Alternate format: [Web html](https://dudek-com.github.io/gvtp/)

# 1 Introduction

Developing renewable energy in an environmentally responsible manner requires mitigation of potentially sensitive species. Herein we characterize the habitat of the Gaviota tarplant ( *Dienandra increscens ssp. villosa*) for the Strauss Wind Energy Site in north Santa Barbara County, California area for the purposes of delineating habitat and informing potential enhancement and restoration efforts.

To describe habitat we built separate models associated with different stages of identification associated with different environmental predictors: 1) **landscape** features describable based on terrain given a digital elevation model (DEM); 2) **biotic** factors related to the plant’s productivity; and 3) **soil** characteristics derived from lab analysis. Each of these stages represents an increasingly detailed view requiring additional levels of effort to assess. This phased approach is meant to reduce the time and expense for facilitating future survey and restoration efforts.

Given these environmental predictors various responses were used to build the statistical relationships based on the binary presence or absence and nine continuous responses related to abundance, percent cover, height, flowering and fruiting. We use the most sophisticated common methods for building species distribution models, namely ***MaxEnt*** for binary presence/absence (Elith et al., 2011; Elith & Leathwick, 2009; Fletcher et al., 2019; Merow et al., 2013; Phillips et al., 2017) and for continous responses ***RandomForest*** (Evans et al., 2011; Kosicki, 2020; Luan et al., 2020; Zhang et al., 2020).

# 2 Methods

The observational data was limited to 40 observations nearly evenly weighted by presence (n=21) and absence (n=19) of Gaviota tarplant (GVTP; *Dienandra increscens ssp. villosa*). The environmental predictors were much more numerous (n=81; Table 2.2) than response terms (n=10) and were classified into 19 landscape, 7 biotic and 55 soil predictor categories ((Table 2.1).

Table 2.1: Biological response variables with short and long names.

|  |  |
| --- | --- |
| Response,short | Response,long |
| g\_cnt | gvtp\_count |
| g\_cov | gvtp\_cover |
| g\_flwr | gvtp\_percent\_flowering |
| g\_fruit | gvtp\_percent\_fruiting |
| g\_head | gvtp\_heads |
| g\_ht | gvtp\_height\_cm\_avg |
| g\_perf | gvtp\_performance |
| g\_pres | gvtp\_presence |
| g\_repro | gvtp\_reproductive\_potential |
| g\_veg | gvtp\_percent\_vegetative |

Table 2.2: Environmental predictor variables by category with short and long names.

|  |  |  |
| --- | --- | --- |
| Category | Predictor,short | Predictor,long |
| biotic | hrb\_ht | herbheight\_avg\_cm |
| biotic | nativ | pct\_native\_cover |
| biotic | nativity | pct\_nativity |
| biotic | nonnativ | pct\_nonnative\_cover |
| biotic | plant | pct\_plant\_cover |
| biotic | wt\_g | dry\_wt\_g |
| biotic | wt\_lbs | dry\_wt\_lbs\_acre |
| landscape | asp | aspect |
| landscape | asp\_cir | aspect\_cir |
| landscape | asp\_cir\_deg | aspect\_cir\_deg |
| landscape | asp\_cir\_e | aspect\_cir\_east |
| landscape | asp\_cir\_n | aspect\_cir\_north |
| landscape | asp\_cir\_rad | aspect\_cir\_rad |
| landscape | asp\_deg | aspect\_deg |
| landscape | asp\_e | aspect\_east |
| landscape | asp\_n | aspect\_north |
| landscape | asp\_rad | aspect\_rad |
| landscape | bare | bare\_grnd\_cover |
| landscape | dstrb | soil\_dist\_cover |
| landscape | dstrb\_cir | soil\_dist\_cover\_cir |
| landscape | elev | elevation |
| landscape | slp\_deg | slope\_degrees |
| landscape | slp\_deg\_cir | slope\_degrees\_cir |
| landscape | slp\_pos | slope\_position |
| landscape | slp\_shp | slope\_shape |
| landscape | slp\_shp\_cir | slope\_shape\_cir |
| soil | ag | silver |
| soil | al | aluminum |
| soil | anion | anion sum |
| soil | as | arsenic |
| soil | b | boron |
| soil | b\_b | boron as B |
| soil | ba | barium |
| soil | ca | calcium |
| soil | ca\_mgl | calcium\_mgl |
| soil | ca\_mil | calcium\_millieq |
| soil | cation | cation sum |
| soil | cd | cadmium |
| soil | cl | chloride |
| soil | cl\_mgl | chloride\_mgl |
| soil | cl\_mil | chloride\_millieq |
| soil | co | cobalt |
| soil | cr | chromium |
| soil | cu | copper |
| soil | fe | iron |
| soil | gyps | est. gypsum requirement-lbs./1000 sq. ft. |
| soil | hg | mercury |
| soil | hlf\_sat | half saturation percentage |
| soil | k | potassium |
| soil | k\_mgl | potassium\_mgl |
| soil | k\_mil | potassium\_millieq |
| soil | li | lithium |
| soil | mg | magnesium |
| soil | mg\_mgl | magnesium\_mgl |
| soil | mg\_mil | magnesium\_millieq |
| soil | mn | manganese |
| soil | mo | molybdenum |
| soil | moist | moisture content of soil |
| soil | n | nitrate |
| soil | n\_mgl | nitrate as N\_mgl |
| soil | n\_mil | nitrate as N\_millieq |
| soil | na | sodium |
| soil | na\_mgl | sodium\_mgl |
| soil | na\_mil | sodium\_millieq |
| soil | ni | nickel |
| soil | organic | organic matter |
| soil | p | phosphorus |
| soil | p\_mgl | phosphorus as P\_mgl |
| soil | p\_mil | phosphorus as P\_millieq |
| soil | pb | lead |
| soil | ph | pH |
| soil | rel\_infil | relative infiltration rate |
| soil | salin | salinity |
| soil | sar | SAR |
| soil | se | selenium |
| soil | soil\_txtr | estimated soil texture |
| soil | sr | strontium |
| soil | su | sulfur |
| soil | su\_mgl | sulfate as S\_mgl |
| soil | su\_mil | sulfate as S\_millieq |
| soil | zn | zinc |

Certain predictor and response terms were created or transformed to elicit meaningful relationships. The reproductive potential response was calculated as the number of GVTP plants multiplied by the average count of heads (). Since aspect is a circular variable (0°— north to 360°— north), it was transformed to radians () to then create variables representing northern () and eastern exposure () (Kvasnes et al., 2018).

Given many more predictors (n=81) than observations (n=40), the first challenge of this analysis was to winnow predictors down to a reasonable subset so as to avoid issues of overfitting and multicollinearity. To do this, we pre-selected predictors (x) across the suite of responses (y) based on individual significant relationships with responses (p.value <= 0.05) for either a simple linear model (y ~ x) or one with an additional quadratic term (y ~ x + x^2) to allow for non-linear niche response (i.e. bell-shaped biological response around some optimal environmental predictor value). Predictors were further filtered for those not autocorrelated (Spearman’s correlation <= 0.7) based on the individual model with the lowest Akaike information criterion (AIC) to produce the most parsimonious model without autocorrelated predictors.

Once a reasonable subset of predictors were chosen per response, data were randomly split into training data (80%) to fit the model and test data (20%) to evaluate model performance. Model performance was based on Accuracy (higher score is better) for the binary response (presence) to compare Maxent against RandomForest models. To compare the continuous response models, a Normalized Root Mean Square Error (NRMSE; lower score is better) was used to allow for model selection between different response terms.

To then select which models to employ for each category (landscape, biotic, soil) requires a value judgement. We propose that the presence response be used at the landscape level as a logical, simple first step in habitat assessment. For this type of binary response data Maxent has been demonstrated to be especially performant with few observations (Fletcher et al., 2019; Hijmans & Elith, 2013). For the subsequent biotic and soil categories, a model based on continuous response could then inform on the reproductive potential of the species amongst within a set of landscape-derived sites.

# 3 Results

Of the many possible combinations between response, predictors and linear model variant (10 \* 81 \* 2 = 1620) a subset of significant, non-correlated predictors was matched for each response, based on binary (i.e. presence/absence; Table 3.1) or continous response (Table 3.2).

Table 3.1: Models with binary biological response (GVTP presence/absence), assessed by highest Accuracy between modeling methods of Maxent and RandomForest.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Response | Predictors | Accuracy, RandomForest | Accuracy, Maxent |
| biotic | g\_pres | hrb\_ht + nonnativ + plant + wt\_g | 0.750 | 1.000 |
| landscape | g\_pres | asp\_cir + bare + dstrb\_cir + slp\_deg\_cir + slp\_pos | 0.875 | 1.000 |
| soil | g\_pres | as + ca\_mgl + fe + mo + ph + rel\_infil + soil\_txtr + sr + zn | 0.625 | 0.875 |

Table 3.2: Models with continuous biological response of GVTP, assessed between RandomForest models with lowest Normalized Root Mean Square Error (NRMSE).

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Response | Predictors | NRMSE, RandomForest |
| biotic | g\_ht | hrb\_ht + nonnativ + plant + wt\_g | 0.5862016 |
| biotic | g\_veg | asp\_cir + bare + dstrb\_cir + slp\_pos + slp\_deg\_cir | 0.8169411 |
| biotic | g\_flwr | hrb\_ht + nonnativ + plant + wt\_g | 0.8355756 |
| biotic | g\_cnt | plant + wt\_g | 0.8866365 |
| biotic | g\_fruit | hrb\_ht + wt\_g | 1.0314135 |
| biotic | g\_cov | nonnativ + plant | 1.0676078 |
| biotic | g\_repro | plant | 7.0209514 |
| landscape | g\_ht | asp + asp\_cir + dstrb\_cir + slp\_deg\_cir + slp\_pos | 0.8115228 |
| landscape | g\_flwr | asp\_cir + dstrb + dstrb\_cir | 0.9162714 |
| landscape | g\_cnt | dstrb\_cir + slp\_shp\_cir | 0.9995098 |
| landscape | g\_cov | asp\_cir + bare + dstrb\_cir + slp\_pos + slp\_deg\_cir | 1.0019437 |
| landscape | g\_fruit | asp\_cir + bare + dstrb\_cir + slp\_pos + slp\_deg\_cir | 1.0431574 |
| landscape | g\_head | asp\_deg + slp\_pos + slp\_shp | 1.4346212 |
| landscape | g\_repro | asp\_cir + bare + dstrb\_cir + slp\_pos + slp\_deg\_cir | 2.8045311 |
| landscape | g\_veg | dstrb + dstrb\_cir + slp\_pos | 3.2193747 |
| soil | g\_ht | as + ca + cation + fe + mo + ph + rel\_infil + zn | 0.7973356 |
| soil | g\_veg | al + li + pb + rel\_infil | 0.9359612 |
| soil | g\_repro | al | 0.9473671 |
| soil | g\_cov | al | 0.9656032 |
| soil | g\_cnt | al + hlf\_sat | 1.0117367 |
| soil | g\_flwr | cation + fe + mo + ph + rel\_infil | 1.0812958 |
| soil | g\_head | al + as + rel\_infil | 1.6892416 |
| soil | g\_fruit | cu + sar | 4.3056746 |

…

# 4 Discussion

…

Although gvtp\_height\_cm\_avg (g\_ht) has the best performance, ecological interpretability favors gvtp\_percent\_vegetative (g\_veg).

Results may also be packaged in such a way that data could be entered in the future at each categorical stage and predictions run within a few simple R functions. These could also be easily folded into a Shiny application with a user interface for uploading the CSVs to run the predictions.

## 4.1 References

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57.

Evans, J. S., Murphy, M. A., Holden, Z. A., & Cushman, S. A. (2011). Modeling Species Distribution and Change Using Random Forest. In C. A. Drew, Y. F. Wiersma, & F. Huettmann (Eds.), *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications* (pp. 139–159). Springer. <https://doi.org/10.1007/978-1-4419-7390-0_8>

Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, *100*(6), e02710. <https://doi.org/10.1002/ecy.2710>

Hijmans, R. J., & Elith, J. (2013). Species distribution modeling with R. *R Package Version 0.8-11*.

Kosicki, J. Z. (2020). Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. *Environmental and Ecological Statistics*, *27*(2), 273–292. <https://doi.org/10.1007/s10651-020-00445-5>

Kvasnes, M. A. J., Pedersen, H. C., & Nilsen, E. B. (2018). Quantifying suitable late summer brood habitats for willow ptarmigan in Norway. *BMC Ecology*, *18*(1). <https://doi.org/10.1186/s12898-018-0196-6>

Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research*, *227*, 105534. <https://doi.org/10.1016/j.fishres.2020.105534>

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species’ distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887–893. <https://doi.org/10.1111/ecog.03049>

Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2020). Improving prediction of rare species’ distribution from community data. *Scientific Reports*, *10*(1), 12230. <https://doi.org/10.1038/s41598-020-69157-x>