

# Incorporating biotic relationships improves species distribution models: Modeling the temporal influence of competition in conspecific nesting birds

Rachel R. Fern<sup>a,\*</sup>, Michael L. Morrison<sup>a</sup>, Hsiao-Hsuan Wang<sup>a</sup>, William E. Grant<sup>a</sup>, Tyler A. Campbell<sup>b</sup>

<sup>a</sup> Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX 77845, USA

<sup>b</sup> East Foundation, San Antonio, TX, 78216, USA

## ARTICLE INFO

### Keywords:

Biotic interactions  
Birds  
Cactus wren  
Competition  
Nesting  
Species distribution model  
Predation  
Verdin

## ABSTRACT

**Aim:** Complex, biotic interactions are notably excluded from species distribution models (SDMs) as they are often difficult to quantify and accommodate in a traditional modeling framework, especially those with a temporal component. The territorial nature of breeding Cactus wren is well-documented and typically involves nest usurping (i.e., destruction) of conspecifics. Due to their similar nesting ecology, breeding Verdin are frequently the target of such behavior and are often forced to move or abandon nests. Using the Verdin/Cactus wren system as a case study, our goal was to evaluate the performance of SDMs that include only environmental predictors with SDMs that also include biotic relationships as predictors.

**Location:** East Foundation's San Antonio Viejo Ranch in south Texas.

**Methods:** We built SDMs (MaxEnt, Boosted Regression Tree [BRT], and Random Forest [RF]) to project Verdin distribution during the early (April through mid-May), peak (mid-May through mid-June), and late (mid-June through mid-July) breeding periods using occurrence data collected during the 2015 and 2016 breeding seasons. We ran parallel analyses using relevant environmental features alone as predictors and then environmental features with observed Cactus wren density.

**Results:** Random Forest (RF) produced the highest predictive performance SDMs for all three breeding periods (AUC = 0.81-0.99; TSS = 0.23-0.73). All models improved in predictive power ( $\Delta$  AUC = 0.01-0.10) and model sensitivity ( $\Delta$  TSS = 0.09-0.66) with the inclusion of Cactus wren density as a predictor of Verdin presence.

**Main conclusions:** Our results indicate that SDM performance is improved by the inclusion of biotic relationships as predictors. Incorporating biotic interactions, as well as their temporal trends, is essential in efforts to monitor or conserve bird species with similar nesting ecologies. Further, modeling algorithms that can accommodate complex, non-linear relationships (e.g., Random Forest) should be preferred in SDM development and application.

## 1. Introduction

Traditionally, species distribution models (SDMs), which statistically associate a species' occurrence with a suite of geospatial predictors, use resource variables, resources that the animal consumes or requires to persist in an area (e.g., shrub density, water availability) to define and project a species' niche and distribution (Austin and Niel, 2011; Elith and Leathwick, 2009). Indirect variables, which are features the animal does not consume or require for persistence but with which it may still interact (e.g., competition or predation), are often not included in SDMs due to the difficulty in identifying the variable

ecologically, quantifying the relationship, or in managing collinearity issues adding an interaction factor to a traditional SDM framework may contribute (Austin and Niel, 2011). However, recent advances in machine learning algorithms (e.g., Random Forest) have enabled us to include indirect variables or biotic relationships, such as competitive exclusion, in SDMs (Miller, 2010). Complex, biotic interactions are notably excluded from SDMs as they are often difficult to quantify and accommodate in a traditional modeling framework, especially those with a temporal component (i.e., an interaction or relationship that changes through time).

In this study we use the ecological relationships between two bird

\* Corresponding author at: Department of Wildlife and Fisheries Sciences, Texas A&M University, WFES Building, 2258 TAMU, 534 John Kimbrough Blvd., College Station, TX, 77843-2258, USA.

E-mail address: [Rachel.Fern@tamu.edu](mailto:Rachel.Fern@tamu.edu) (R.R. Fern).

<https://doi.org/10.1016/j.ecolmodel.2019.108743>

Received 16 April 2019; Received in revised form 5 July 2019; Accepted 7 July 2019

0304-3800/ © 2019 Published by Elsevier B.V.

species as a case study of the potential value of incorporating biotic interactions into SDMs. Competition between sympatric species often include contest for limited resources (e.g., territory, food, mates). In birds, availability of nesting space or materials presents a unique competitive challenge in that the degree of contest between species has a temporal aspect, namely the breeding seasons. For example, competition for nest-sites is often observed in sympatric bird species with similar nesting ecology that would otherwise partition resources the remainder of the year (Martin, 1993; MacArthur, 1958). Inferior competitors are then frequently pushed to alternative nest-sites that may be more vulnerable to predation or exposure (Newton, 1994).

One such example is the relationship between breeding Cactus wren *Campylorhynchus brunneicapillus*, a large (32–47 g), territorial wren native to the desert southwestern United States, and Verdin *Auriparus flaviceps*, a small (5–8 g) penduline tit native to the southwestern United States and northern Mexico (Anderson and Anderson, 1973; Lockwood and Freeman, 2004; Williamson, 2000). Most of the year, the disparity in size and general morphology of these two species allows for territorial and food resource partitioning (Anderson and Anderson, 1973; Wise-Gervais, 2005). However, both the Cactus wren and Verdin prefer areas of dense Cholla cacti (*Opuntia spp.*) for nesting but will also utilize Palo Verde (*Cercidium spp.*), large mottes of mixed cacti, and other spiny plants (Anderson and Anderson, 1973; Wise-Gervais, 2005). The territorial nature of breeding Cactus wren is well-documented and typically involves nest usurping (i.e., destruction), with the Verdin frequently the target of such behavior and are often forced to retreat to alternative nest-sites (McGee, 1985; Simons and Simons, 1990).

Our main objective was to evaluate the performance of SDMs that include only environmental predictors with SDMs that also include biotic relationships as predictors. More specifically, as a proof of concept, we evaluated the ability of SDMs of Verdin that included only environmental predictors with SDMs that also included the density of Cactus wren as a predictor. We also evaluated the ability of SDMs incorporating Cactus wren to detect differences in the relative influence of this biotic predictor among early, peak, and late portions of the Verdin breeding season.

## 2. Methods

### 2.1. Study site

We conducted our study on the East Foundation's 61,000-ha San Antonio Viejo Ranch (SAV), located approximately 25 km south of Hebronville, Texas in Jim Hogg and Starr counties. SAV is located within the South Texas Plains ecoregion and is one of six properties of the East Foundation that are managed as a living laboratory to support wildlife conservation and other public benefits of ranching and private land stewardship. Vegetation composition and structure within our study area is characteristic of this ecoregion and consists of a mosaic of grassland and thornscrub. SAV is representative of south Texas rangeland ecosystems and encompasses the Coastal Sand Plain and Texas-Tamulipan Thornscrub ecoregions. Low-growing woody plants, dense shrubs (*Prosopis glandulosa*, *Acacia greggii*, *Celtis ehrenbergiana*, *Colubrina texensis*, *Aloysia gratissima*, *Lantana urticoides*), and cacti (*Opuntia engelmannii* var. *lindheimeri*, *Opuntia leptocaulis*) dominate the vegetation in this area. Mean annual temperature within our study site is 22.6 °C and mean annual precipitation is 502.5 mm (PRISM Climate Group, 2018). Mean monthly temperature during our study period (April–July 2015 and 2016) was 27.2 °C with a maximum daily high of 32.8 °C (PRISM Climate Group, 2018). Mean monthly precipitation during our study period was 77.1 mm with a maximum daily high of 93.1 mm (PRISM Climate Group, 2018).

### 2.2. Environmental predictors

We used canopy height, shrub density, water proximity, grass *spp.*

coverage, and cacti *spp.* coverage recorded from ground surveys as well as local topographic relief and an Optimized Soil Adjusted Vegetation Index (OSAVI) calculated using remotely sensed imagery in 2015 and 2016 as environmental predictors in SDMs. Specific workflow and processing details for each environmental predictor can be found in Appendix A.

### 2.3. Bird occurrence data

Avian point counts consisted of 25 12-point transects arranged in a stratified-random design across SAV, stratified by vegetation type. Each transect was surveyed 3 times throughout the breeding season: first visit between April and mid-May, second visit between mid-May and mid-June, and third visit between mid-June and mid-July. We used point count data collected from April to July 2015 and 2016 to build baseline SDMs. Each point was located 400-m apart, and 2 observers recorded visual and auditory occurrences of birds within 200-m of each point simultaneously yet independently. We used a traditional framework in which each occurrence was counted as a 'presence' record at each point, omitting the duplicate records from the double observer design, and resampling the data by a 200-m cell-size. This granted us a finer spatial resolution of the data set to thoroughly investigate the influence of conspecific (Cactus wren) density on the presence of Verdin. Cactus wren density was calculated using observed abundance at each transect point divided by the total space surveyed (200-m<sup>2</sup>). We used the occurrence of Verdin and the calculated density of Cactus wren during the breeding seasons of 2015 and 2016 for SDMs.

### 2.4. Data processing and analysis

We exported the rasters of all predictors as GeoTIFFs and read these into the R statistical language as raster layers (R Core Team, 2013). All layers were stacked to create the occurrence predictor stack for SDMs. Parallel analyses were also performed by time period within the breeding season: early (April through mid-May), peak (mid-May through mid-June), and late (mid-June through mid-July) to investigate temporal changes in Cactus wren influence on Verdin distribution. Bird occurrence data were subset by season. We imported occurrence data for Verdin into R and used the predictor raster stack to build SDMs using three different algorithms: MaxEnt, Boosted Regression Tree (BRT), and Random Forest (RF). Table 1 outlines the basic mathematical approach of each modeling algorithm and provides a comparison of the advantages of each model in the occupancy framework. We generated 'background data' to produce the non-presence class required by the logistic models. Background data do not attempt to guess at absence locations, but instead are used to characterize the study region (Phillips and Elith, 2011; Phillips et al., 2009; Ward et al., 2009). Background data represent the environmental domain of the study and are independent of occurrence data while presence data establish the conditions under which a species is more likely to be present than a null, or completely random, model would predict. After building baseline SDMs, we added the raster representing Cactus wren density to the occurrence predictor raster stack and re-ran the models to assess any improvement or degradation in the predictive performance and sensitivity of each algorithm. For each breeding period, we also calculated relative influence and significance of each predictor using a mixed-model approach. We calculated relative influence of each predictor on Verdin presence using a BRT analysis and the significance of highly influential variables through logit generalized linear regression analyses. The BRT has the unique advantage to accommodate collinearity among predictors and fit complex nonlinear relationships between response and predictor variables making it ideal for determining the relative contribution of each predictor (Elith et al., 2008; Franklin, 2010). Generalized linear models, based on generalized multiple linear regression, also accommodate non-linear relationships through use of the "link" function in which predictors can be transformed based on

**Table 1**

Comparison of mathematical approach for each modeling algorithm being used to project species distributions in this study. Data requirements and advantages are also listed.

Model	Data type	Approach/mechanism	Advantages
MaxEnt	Presence only	A machine-learning algorithm based on the principle from statistical mechanics and information theory that states that the probability distribution with maximum entropy is the best approximation of an unknown distribution (Phillips et al., 2006).	Recent investigations have shown the MaxEnt algorithm to be mathematically identical to that of the GLM (Poisson distribution) (Renner and Warton, 2013). Its unique ability to accept environmental gradients as part of the prediction process make its application to ecological niche modeling ideal (Evangelista et al., 2009; Saatchi et al., 2008).
Boosted Regression Tree (BRT)	Presence/absence	An ensemble, regression-based method that combines the strengths of two commonly used algorithms: regression trees (models that define the response to predictors using binary splits) and boosting (a method for combining multiple simple models to improve performance). An initial regression tree is fitted and iteratively improved upon in a forward stagewise manner (boosting) by minimizing the variation in the response not explained by the model at each iteration.	This approach can easily accommodate different types of predictor variables, missing data, and outliers as well as fit complex nonlinear relationships automatically handling collinearity between predictor variables. BRT interpretations can be easily summarized to provide powerful ecological insight (Franklin, 2010).
Random Forest (RF)	Presence/absence	An ensemble machine-learning method in which a large number (500-2000) of decision trees are grown with subsets of the data (e.g., species occurrences) containing a random subset of candidate predictor variables (Breiman, 2001). Each tree votes for a binary outcome and the resulting predictions are averaged.	This method makes no assumptions on data distribution and instead uses bootstrap aggregation to subsample the given data. This approach has been shown to have higher prediction accuracy than ordinary decision trees in SDM and other applications. (Gislason et al., 2006; Prasad et al., 2006).

response data distribution (Austin and Cunningham, 1981; Franklin, 2010; Margules et al., 1987). We used the logit distribution (binary response or presence/absence) to determine significance of each predictor.

Prior to building SDMs, we performed preliminary analyses to ensure only predictors that added to the explanatory power of the models and did not add to the overall deviance were used in each SDM. This included the use of a priori Gradient Boosting Machine (GBM) analyses and step-wise regression variable dropping and selection for each model and time period. We also performed time-fixed effects (FE) regression analysis to identify any significant effects between years in occurrence response to predictor variables. There was no significant effect in Verdin occurrence response to predictor variables between years ( $p = 0.79$ ), thus eliminating the need to separate data per year or accommodate for time-FE in SDMs. Occurrence data were pooled for the 2015 and 2016 breeding seasons.

### 2.5. Model evaluation

We evaluated performance of each model using the Area Under the Receiver Operator Curve (AUROC or AUC) and true sensitivity statistic (TSS). The AUC (range from 0 to 1) is a measure of rank-correlation. In unbiased data, a higher AUC value indicates that areas with high predicted suitability values tend to be sites of known presence (Phillips et al., 2006). The TSS is an approach based on maximizing the sum of sensitivity and specificity independent of species prevalence (Liu et al., 2013). Many distributional model evaluation approaches (e.g., kappa) are threshold-dependent; a value above a user-set threshold indicates a prediction of presence and a value below the threshold indicates absence. However, different models assign different weight to false absences or false presences making it hard to compare models directly. The TSS is considered an alternative to the traditionally used kappa to assess model performance, since it has the advantage of being threshold and prevalence independent.

## 3. Results

We recorded a total of 981 occurrences of Verdin throughout the 2015 and 2016 breeding seasons: 351 during the early breeding period (April through mid-May), 322 during the peak breeding period (mid-May through mid-June), and 308 during the late breeding period (mid-June through mid-July). We also recorded a total of 788 occurrences of Cactus wren throughout the 2015 and 2016 breeding seasons: 332

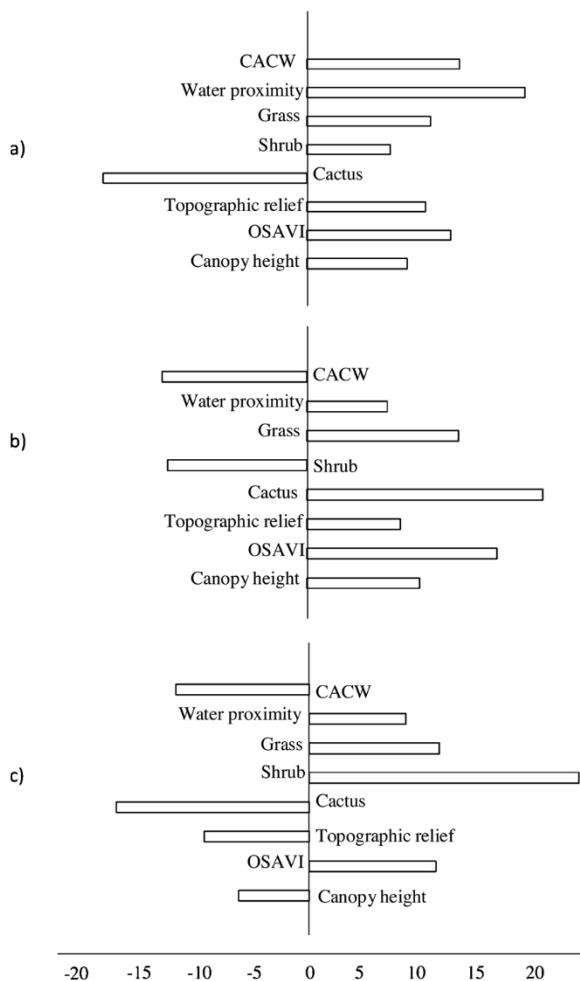
during the early breeding period, 278 during the peak breeding period, and 178 during the late breeding period. Estimated Cactus wren densities ranged from 0.06 to 2.9 individuals per 200-m<sup>2</sup>. Of the three algorithms used, Random Forest (RF) produced the highest predictive performance SDMs for all three breeding periods (Table 2). Boosted Regression Tree (BRT) produced the lowest performing model overall for the early breeding period, both, in terms of predictive power (AUC = 0.60, 0.63; with and without Cactus wren density included a predictor, respectively) and model sensitivity (TSS = -0.02, 0.15; with and without Cactus wren density included as predictor, respectively).

All models improved in both predictive power ( $\Delta$  AUC = 0.01-0.10) and model sensitivity ( $\Delta$  TSS = 0.09-0.66) with the inclusion of Cactus wren density as a predictor of Verdin presence (Table 2). However, magnitude of improvement in model performance varied by breeding period. During the early breeding period (April through mid-May), SDMs that included Cactus wren density as a predictor performed only slightly better than those that included environmental features alone (MaxEnt:  $\Delta$  AUC = +0.02,  $\Delta$  TSS = +0.09; BRT:  $\Delta$  AUC = +0.03,  $\Delta$  TSS = +0.17; RF:  $\Delta$  AUC = +0.01,  $\Delta$  TSS = +0.24). SDMs that included Cactus wren density as a predictor for the peak breeding season (mid-May through mid-June), as opposed to those including environmental features only, produced larger increases in model performance, relative to the early breeding period (MaxEnt:  $\Delta$  AUC = +0.09,  $\Delta$  TSS

**Table 2**

Results of species distribution model (SDM) performance for MaxEnt, Boosted Regression Tree (BRT), and Random Forest (RF) algorithms in predicting occurrence of Verdin *Auriparus flaviceps* on East Foundation's San Antonio Viejo Ranch (SAV) during the early (April through mid-May), peak (mid-May through mid-June), and late (mid-June through mid-July) breeding seasons of 2015–2016. Model performance metrics (area under curve [AUC] and true sensitivity statistic [TSS]) are compared for SDMs using environmental predictors only and environmental predictors stacked with a raster representing Cactus wren *Campylorhynchus brunneicapillus* density (denoted by '+').

		MaxEnt		BRT		RF	
		AUC	TSS	AUC	TSS	AUC	TSS
Early	-	0.74	0.40	0.60	-0.02	0.81	0.23
	+	0.76	0.49	0.63	0.15	0.82	0.47
Peak	-	0.80	0.14	0.89	0.48	0.89	0.59
	+	0.89	0.34	0.95	0.73	0.99	0.67
Late	-	0.81	0.27	0.73	0.31	0.88	0.59
	+	0.86	0.93	0.77	0.46	0.98	0.73



**Fig. 1.** Relative influence, as calculated by boosted regression tree analysis (BRT), of environmental features and Cactus wren *Campylorhynchus brunneicapillus* (CACW) density on the presence of Verdin *Auriparus flaviceps* on East Foundation's San Antonio Viejo Ranch during the early (a; April through mid-May), peak (b; mid-May through mid-June), and late (c; mid-June through mid-July) 2015–2016 breeding seasons.

= +0.20; BRT:  $\Delta$  AUC = +0.06,  $\Delta$  TSS = +0.25; RF:  $\Delta$  AUC = +0.10,  $\Delta$  TSS = +0.08). SDMs built for the late breeding period (mid-June through mid-July) produced similar, yet less pronounced improvements in performance for models that included Cactus wren density as a predictor as opposed to environmental features alone (MaxEnt:  $\Delta$  AUC = +0.05,  $\Delta$  TSS = +0.66; BRT:  $\Delta$  AUC = +0.04,  $\Delta$  TSS = +0.15; RF:  $\Delta$  AUC = +0.10,  $\Delta$  TSS = +0.14).

Water proximity was the most influential (+) feature in predicting Verdin presence during the early breeding period (Fig. 1). Verdin presence during the peak breeding season was most influenced by Cactus wren density (-) and green biomass (+), represented by the OSAVI. Cactus wren density (-) and green biomass (+) remained influential in the late breeding period. However, shrub density (+) and cactus *spp.* coverage (-) also became influential predictors for Verdin presence.

#### 4. Discussion

Our findings indicate inclusion of a biotic relationship into a traditional SDM framework improves model predictive power and model sensitivity. Incorporating these biotic interactions, as well as their temporal trends, is essential in efforts to monitor or conserve bird species with similar nesting ecologies. Ecologists should not only consider the environmental requirements for species persistence, but also

the presence of conspecifics with which they are known to interact during various life history stages (e.g., fledging periods, nest-site selection or initiation). Modeling algorithms that can accommodate complex, non-linear relationships (e.g., Random Forest) should be preferred in SDM development and application. Random Forest routinely outperforms other machine-learning and linear algorithms, both, in our study and others involving non-normal data distribution and complex predictor interactions (Breiman, 2001; Mi et al., 2014; Prasad et al., 2006).

Although our objective was not to thoroughly examine interactions between these two bird species, our results demonstrate how SDMs can be used to investigate potential competitive interactions between animal species. For example, Cactus wren density had a significant, negative influence on the distribution of Verdin during the peak (mid-May through mid-June) and late (mid-June through mid-July) breeding season. However, Cactus wren density was not a negatively correlated predictor for Verdin presence during the early breeding season (April through mid-May). This may have biologically meaningful implications for understanding how these two conspecifics interact during nest-site selection. More specifically, this suggests that optimal nest-site selection may not necessarily be the source of the observed competition pressure. Observational studies suggest that Cactus wren usurp conspecific nests in an effort to maximize predator-free nest space (McGee, 1985; Simons and Simons, 1990). Our findings support this assertion as the potential source of competitive pressure since Verdin presence was not negatively affected by Cactus wren density until after the initial breeding period. Further, predation is the primary cause of nest failure in birds so it is reasonable to expect species able to minimize predation pressure will have a substantial competitive advantage (Davis, 2017; Newton, 1994). It is imperative we include such biotic relationships, as well as their temporal components, in the modelling framework for accurate SDMs.

#### Biosketch

The focus of our research is improving distribution models by incorporating complex, biotic interactions through the use of various algorithms, data transformation and generation, and remotely sensed imagery data. We continue to investigate new ways to include temporal influence and biotic relationships into a traditional predictive modeling framework to create more accurate and inclusive species distribution models.

#### Data accessibility

All environmental GIS layers and remotely sensed imagery generated for this study are available as raster grids from the joint National Aeronautics and Space Administration (NASA) and U.S. Geological Survey (USGS) database for Landsat imagery data at: [landsat.usgs.gov](https://landsat.usgs.gov). Climate data generated for this study are available as 'band interleaved by line' formatted imagery, convertible for raster grids, from the PRISM Climate Group database at: [prism.oregonstate.edu](https://prism.oregonstate.edu).

#### Acknowledgment

We offer our gratitude to the East Foundation for their generous funding and access to the San Antonio Viejo Ranch, as well as the many field technicians that aided in the ground surveys. This is manuscript number 037 of the East Foundation.

#### Appendix A

We collected vegetation composition and structure data from 141 permanent 20-m transects in October 2016. We allocated transects proportional to the area of ecological sites that occur in each pasture using stratified sampling resulting in 12–16 transects per pasture

(Bonham, 2013). We marked each transect start and collected data in a random, predetermined direction (N, S, E, W). On each transect we sampled 5, 20 × 50 cm quadrats (5 m spacing) randomly placed at either 0, 0.5, 1, 1.5, 2, or 2.5 m from the left side of the tape and facing away from the transect start, visually recording percent cover of woody and herbaceous (later classified by grass *spp.*) in each quadrat.

We also documented woody canopy cover along each of the 20 m transects by visually recording the amount of the ground (in centimeters) covered by woody plant materials (leaves and branches) and succulent (cacti) that intercepted the line transect by species (Canfield, 1941; Higgins et al., 1996). If a gap in the canopy exceeded 0.5 m for an individual, we recorded separate cover measurements. We calculated percent canopy cover by summing the intercept measurements for an individual species, dividing by total line length and converting to a cover percentage. We calculated total percent cover by adding cover percentages for all species, which sometimes exceeded 100% when overlapping canopies by different species were recorded (Coulloudon et al., 1999). Locations of water sources (e.g., livestock wells) within the study site were provided by the East Foundation. To calculate water proximity, we gridded the spatial extent of the SAV Ranch into a fishnet (30-m<sup>2</sup> resolution). We performed a proximity analysis on each pixel centroid using the Near tool in ArcMap 10.5 to determine distance of each centroid to location of nearest water source, usually a livestock well and holding tank as no natural surface water exists within the study site, and very little exists on the Coastal Sand Plain region of Texas as a whole (Snelgrove et al., 2013). We made considerations for seasonality as not all groundwater pumps are operational year-round on large south Texas cattle ranches and ensured only those wells known to be active during the summer of 2015 and 2016 were used in the analysis.

We imported values for each predictor (canopy height, shrub density, water proximity, grass *spp.* coverage, and cacti *spp.* coverage) into ArcMap 10.5 and used Kriging interpolation to minimize spatial sampling bias and create continuous surface layers of environmental predictor values. Kriging is a geostatistical method through which interpolated values are modeled by a Gaussian process governed by covariances. This method of spatial interpolation estimates a continuous surface of values directly based on values at surrounding points weighted according to spatial covariance (van Beers and Kleijnen, 2004). The Kriging interpolation algorithm is optimal for most eco-spatial modeling because it produces an unbiased prediction and calculates the spatial distribution of uncertainty allowing for an accurate estimate of error at any particular point (Mahmoudabadi and Briggs, 2016). We then calculated the mean values for each predictor within the 2015 and 2016 study periods using the raster algebra tool in ArcMap 10.5.

We acquired one Landsat 8-OLI tile (< 6% cloud cover) that encompassed the study area (courtesy of U.S. Geological Survey) for each year (2015 and 2016) and processed this in ENVI 5.1 (NASA Landsat Program, 2015; 2016). We corrected for atmospheric conditions and converted the original image format of Digital Numbers (DN) to radiance and then surface reflectance. We first resized the images to the rectangular extent of the SAV and then extracted by the study area mask in ESRI ArcGIS ArcMap 10.5. We then spatially subset each extracted image by bands 2–5 corresponding to Landsat 8-OLI band designations: blue, green, red, and NIR. Bands were stacked and the OSAVI was calculated using the band math tool in ENVI 5.1 for each image. This index for LAI follows the standard formula [(NIR-Red)/(NIR + Red + 0.16)] and uses a reflectance constant of 0.16 to adjust for high background reflectance (e.g., areas with sparse vegetation and high soil reflectance) (Rondeaux et al., 1996). In south Texas, specifically, this vegetation index outperforms other, more common vegetation indices (e.g., Normalized Difference Vegetation Index [NDVI]) in, both, overall image classification accuracy and herbaceous coverage estimations (Fern et al., 2018). We calculated the mean OSAVI values for the 2015 and 2016 study periods using the raster algebra tool in ArcMap 10.5.

## References

- Anderson, A.H., Anderson, A., 1973. The Cactus Wren. University of Arizona Press, Tucson, AZ.
- Austin, M.P., Cunningham, R.B., 1981. Observational analysis of environmental gradients. *Pro. Ecol. Society Aust.* 11, 109–119.
- Austin, M.P., Niel, K.P., Van, 2011. Improving species distribution models for climate change studies: variable selection and scale. *J. Biogeogr.* 38, 1–8.
- van Beers, W.C.M., Kleijnen, J.P.C., 2004. Kriging Interpolation in Simulation: a Survey Vol. 1 121.
- Bonham, C.D., 2013. Measurements for Terrestrial Vegetation, 2nd ed. John Wiley & Sons, Chichester, West Sussex UK.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Canfield, R.H., 1941. Application of the line interception method in sampling range vegetation. *J. For.* 39, 388–394.
- Coulloudon, B., Eshelman, K., Gianola, J., Habich, N., Hughes, L., Johnson, C., Pellant, M., Podborny, P., Rasmussen, A., Robles, B., 1999. Sampling Vegetation Attributes. BLM Technical Reference 1734.
- Davis, H.T., 2017. Influence of Vegetative Characteristics on Predation and Predator Assemblage of Bird Nests. Texas A&M University M.Sc. Thesis.
- Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813.
- Evangelista, P.H., Stohlgren, T.J., Morissette, J.T., Kumar, S., 2009. Mapping invasive tamarisk (Tamarix): a comparison of single-scene and time-series analyses of remotely sensed data. *Remote Sens.* 1, 519–533.
- Fern, R.R., Foxley, E.A., Bruno, A., Morrison, M.L., 2018. Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland. *Ecol. Indic.* 94, 16–21.
- Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge, United Kingdom.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. *Pattern Recognit. Lett.* 27, 294–300.
- Higgins, K.F., Oldemeyer, J.L., Jenkins, K.J., Clamby, G.K., Harlow, R.F., 1996. Vegetation sampling and measurement. *Research and management techniques for wildlife and habitats* 5, 567–591.
- Liu, C., White, M., Newell, G., Pearson, R., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* 40, 778–789.
- Lockwood, M.W., Freeman, B., 2004. The TOS Handbook of Texas Birds. Texas A&M University Press, College Station, TX.
- MacArthur, R.H., 1958. Population Ecology of Some Warblers of Northeastern Coniferous Forests. *Ecology* 39, 599–619.
- Mahmoudabadi, H., Briggs, G., 2016. Directional kriging implementation for gridded data interpolation and comparative study with common methods. AGU Fall Meeting Abstracts.
- Margules, C.R., Nicholls, A.O., Austin, M.P., 1987. Diversity of *Eucalyptus* species predicted by a multi-variable environmental gradient. *Oecologia* 71, 229–232.
- Martin, T.E., 1993. Nest Predation and Nest Sites: new perspectives on old patterns. *BioScience* 43, 523–532.
- McGee, M., 1985. Interspecific Nest Influence: the Influence of Cactus wrens (*Campylorhynchus brunneicapillus*) on Verdin (*Auriparus flaviceps*) Nest Site Selection (Mexico, Arizona). University of Arizona M.Sc. Thesis.
- Mi, C., Huettmann, F., Guo, Y., 2014. Obtaining the best possible predictions of habitat selection for wintering Great Bustards in Cangzhou, Hebei Province with rapid machine learning analysis. *Chinese Sci. Bull.* 59, 4323–4331.
- Miller, J., 2010. Species distribution modeling. *Geogr. Compass* 4, 490–509.
- NASA Landsat Program, 2015. LANDSAT 8 OLI/TIRS Collection 1 – Path:27 Row: 41. Scene LC08\_L1TP\_027041\_20150720\_20170310\_01\_T1. USGS, Sioux Falls (07/20/2015).
- NASA Landsat Program, 2016. LANDSAT 8 OLI/TIRS Collection 1 – Path:27 Row: 41. Scene ID:LC08\_L1TP\_027041\_20160706\_20170222\_01\_T1. USGS, Sioux Falls (07/06/2016).
- Newton, I., 1994. The role of nest sites in limiting the number of hole nesting birds: a review. *Biol. Conserv.* 70, 265–276.
- Phillips, S., Elith, J., 2011. Logistic Methods for Resource Selection Functions and Presence-Only Species Distribution Models. *Association for the Advancement of Artificial Intelligence*, pp. 1384–1389.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190, 231–259.
- Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection Bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.
- PRISM Climate Group, 2018. Oregon State University. created 4 Feb 2018. <http://prism.oregonstate.edu>.
- R Core Team, 2013. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Renner, I.W., Warton, D.I., 2013. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* 69, 274–281.
- Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55, 95–107.
- Saatchi, S., Buermann, W., Ter Steege, H., Mori, S., Smith, T.B., 2008. Modeling

- distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sens. Environ.* 112, 2000–2017.
- Simons, L.S., Simons, L.H., 1990. Experimental studies of nest-destroying behavior by Cactus wrens. *Condor* 92, 855–860.
- Snelgrove, A., Dube, A., Skow, K., Engeling, A., 2013. East Wildlife Foundation Atlas. Texas A&M Institute of Renewable Natural Resources, College Station, Texas, USA.
- Ward, G., Hastie, T., Barry, S., Elith, J., Leathwick, J.R., 2009. Presence-only data and the EM algorithm. *Biometrics* 65, 554–563.
- Williamson, S.L., 2000. Blue-throated hummingbird (*Lampornis clemenciae*). In: Poole, A., Gill, F. (Eds.), *The Birds of North America* (No. 531). The Birds of North America, Inc, Philadelphia, PA.
- Wise-Gervais, C., 2005. Verdin (*Auriparus flaviceps*). In: Corman, T.E., Wise-Gervais, C. (Eds.), *Arizona Breeding Bird Atlas*. University of New Mexico Press, Albuquerque, NM.