RESEARCH   **Open Access**

Modeling the influence of livestock grazing pressure on grassland bird distributions

Rachel R. Fern[1]* , Michael L. Morrison[1], William E. Grant[1], Hsiaohsuan Wang[1] and Tyler A. Campbell[2]

Abstract

Background: The influence of vegetative changes due to livestock grazing on grassland birds is well-recognized because these birds are heavily influenced by vegetative structure. Traditionally, species distribution models (SDMs) use direct variables, resources that the animal consumes or requires to persist in an area (e.g., water) to define and project a species' niche and distribution. Indirect variables, which are features the animal does not consume or require for persistence but with which it may still interact, are often excluded. Our objective was to improve the traditional SDMs projecting the distribution of three summer resident South Texas grassland birds (Northern Bobwhite *Colinus virginianus*, Eastern Meadowlark *Sturnella magna*, and Cassin's Sparrow *Peucaea cassinii*) by incorporating livestock grazing pressure, an indirect variable, into five SDM algorithms: BioClim, generalized linear model, MaxEnt, boosted regression tree, and random forest. We collected data from the Coloraditas Grazing Research and Demonstration Area (CGRDA), a 7684-ha area located on the San Antonio Viejo Ranch (SAV) in South Texas. We used several relevant environmental characteristics to build SDMs and compared model performance (AUC and TSS) with and without grazing pressure as an indirect variable.

Results: Machine learning models (MaxEnt and random forest) had the highest predictive performance for all species, with random forest being the most consistent for each analysis. BioClim and generalized linear model remained constant or only marginally improved with the addition of the grazing pressure.

Conclusions: Our findings suggest that model selection for SDM should include consideration of species prevalence, and machine-learning algorithms should be preferred when the target species is of low or unknown prevalence. Further, livestock grazing has measurable influence on grassland bird species' distributions and should be included in SDMs as an indirect variable in addition to associated vegetative changes.

Keywords: Birds, Grassland, Grazing, Livestock, Species distribution model, Texas

Background

Domestic livestock is recognized ecosystem engineers in semi-arid rangelands, where they directly and indirectly alter the availability of resources to a wide range of grassland-associated organisms (Derner et al. 2009). Several studies cite the influence of vegetative changes due to livestock grazing on breeding grassland birds since this species group is heavily influenced by vegetative structure (Askins et al. 2007; Brennan and Kuvlesky Jr 2005; Fuhlendorf et al. 2006; Jansen et al. 1999). Grazing pressure has historically been difficult to quantify due to variable plant responses to grazing and movements of livestock within pastures (Landsberg and Crowley 2004). However, a review of the effects of water-place distribution on rangelands suggested that distances from water sources (e.g., livestock tanks and troughs) can provide valuable context for interpreting changes in grazed landscapes particularly in areas remote from water sources (James et al. 1999; Landsberg and Crowley 2004; Ludwig et al. 2000). In South Texas, water sources are scarce. Specifically, the Coastal Sand Plain region of Texas has

* Correspondence: RachelRFern@gmail.com
[1]Department of Wildlife and Fisheries Sciences, Texas A&M University, WFES Building, 2258 TAMU, 534 John Kimbrough Blvd, College Station, TX 77843-2258, USA
Full list of author information is available at the end of the article

no natural permanent bodies of freshwater making live-stock wells and holding tanks supplied by active ranching operations the only water source for domestic livestock and, thus, it is reasonable to expect water points to have a substantial impact on the distribution and intensity of localized grazing (Fulbright et al. 1990; Snelgrove et al. 2013).

Grassland bird populations have experienced precipitous declines on a continental scale over the last few decades (Brennan and Kuvlesky Jr 2005; Nocera and Koslowsky 2011). Although agricultural and livestock operations dominate the South Texas landscape, wildlife-related recreation has become increasingly important to landowners because of the associated economic value (Dodd 2009; TPWD 2016). For example, landowners can anticipate an average gross profit of $4.69 per hectare for a deer or exotic ungulate hunting lease and can expect an average gross profit of $20.99 per hectare for a quail (e.g., Northern Bobwhite *Colinus virginianus* and Scaled Quail *Callipepla squamata*) hunting lease (TPWD 2017). This area also provides crucial resources for other migratory and resident grassland birds (e.g., Cassin's Sparrow *Aimophila cassinii*, Grasshopper Sparrow *Ammodramus savannarum*, and Dicksissel *Spiza americana*) that have declined throughout their ranges due to land use and climate change since 1966 (Brennan and Kuvlesky Jr 2005; Knopf 1994). It is essential we advance our understanding of how grassland birds are affected by their environment, inclusive of both their requirements to persist (i.e., resources), and how they interact with environmental features or biotic influences.
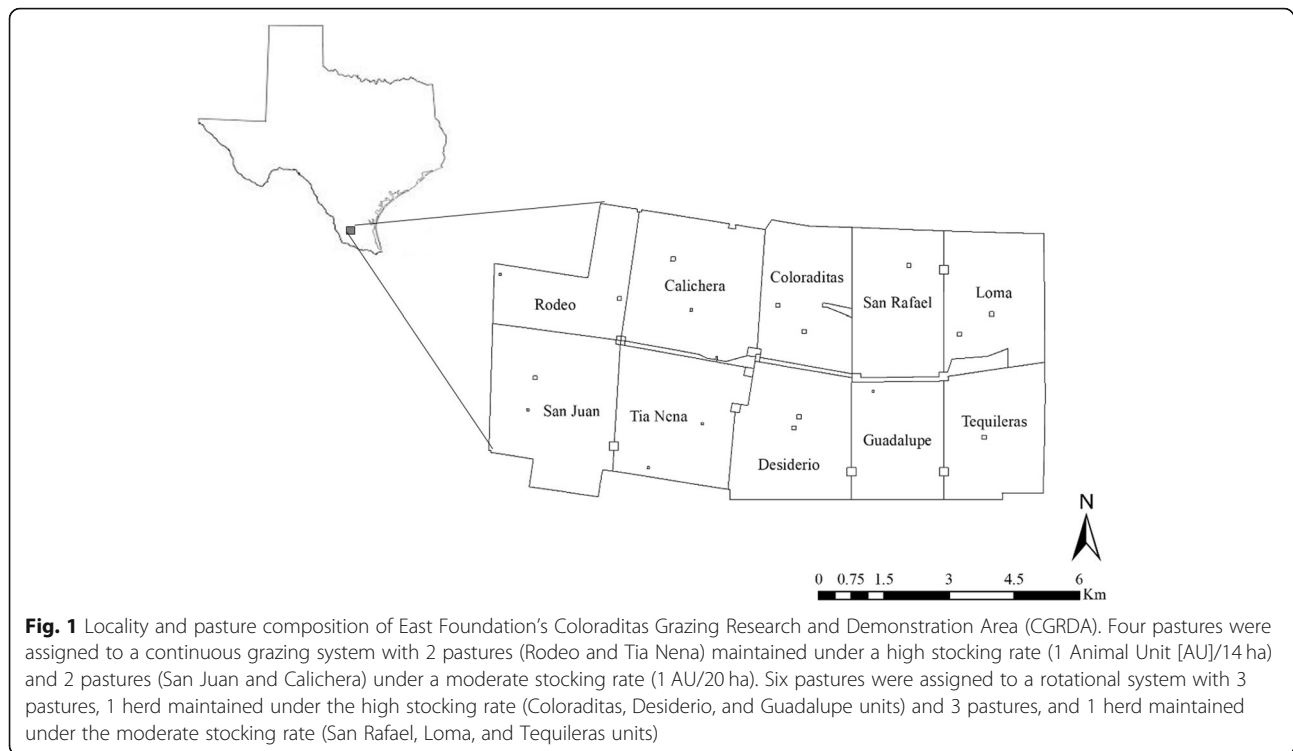
Traditionally, species distribution models (SDMs), which statistically associate a species' occurrence with a suite of geospatial predictors, use direct variables, resources that the animal consumes or requires to persist in an area (e.g., shrub density and water availability) to define and project a species' niche and distribution (Austin and Van 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 and commensalism), are often excluded 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 (Atauchi et al. 2018; Leach et al. 2016; Austin and Van Niel 2011). However, recent advances in machine learning algorithms (e.g., random forest) have enabled us to include indirect variables, such as grazing pressure, in SDMs that may have more complicated relationships with the distribution of the target species than traditional resource variables (Miller 2010).

Our objective was to (1) improve traditional SDMs projecting the distribution of three summer resident South Texas grassland birds (Northern Bobwhite *Colinus virginianus*, Eastern Meadowlark *Sturnella magna*, and Cassin's Sparrow *Peucaea cassinii*) by incorporating live-stock grazing pressure, an indirect variable, and (2) interpret the possible effect of grazing pressure on bird distribution per each SDM approach. We used a novel approach to spatially quantify localized grazing pressure to include this variable using five SDM algorithms: Bio-Clim, generalized linear model, MaxEnt, boosted regression tree, and random forest. Our approach serves as a valuable tool for rangeland managers when the management goal is to promote sustainable livestock grazing and recreational wildlife harvest, while maintaining viable nongame species.

Methods

Study site

The Coloraditas Grazing Research and Demonstration Area (CGRDA) is a 7684-ha area located on the 60,000-ha San Antonio Viejo Ranch (SAV) approximately 25 km south of Hebbronville, Texas, in Jim Hogg and Starr counties (Fig. 1). SAV is located within the South Texas Plains ecoregion and is managed predominantly as a cow-calf operation. Mean annual temperature within the study site is 22.6 °C, and mean annual precipitation is 502.5 mm (PRISM Climate Group 2018). SAV 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. The CGRDA is the representative of South Texas rangeland ecosystems and encompasses the Coastal Sand Plain and Texas-Tamulipan Thronscrub 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. The CGRDA is comprised of 10 pastures each assigned to 1 of 4 grazing systems (Fig. 1). Four pastures were assigned to a continuous grazing system with 2 pastures (Rodeo and Tia Nena) maintained under a high stocking rate (1 Animal Unit [AU] /14 ha) and 2 pastures (San Juan and Calichera) under a moderate stocking rate (1 AU/20 ha). Six pastures were assigned to a rotational system with 3 pastures, 1 herd maintained under the high stocking rate (Coloraditas, Desiderio, and Guadalupe units) and 3 pastures, and 1 herd maintained under the moderate stocking rate (San Rafael, Loma, and Tequileras units). Grazing was deferred on all pastures for 2 years prior to the onset of livestock grazing in December 2015.

Fig. 1 Locality and pasture composition of East Foundation's Coloraditas Grazing Research and Demonstration Area (CGRDA). Four pastures were assigned to a continuous grazing system with 2 pastures (Rodeo and Tia Nena) maintained under a high stocking rate (1 Animal Unit [AU]/14 ha) and 2 pastures (San Juan and Calichera) under a moderate stocking rate (1 AU/20 ha). Six pastures were assigned to a rotational system with 3 pastures, 1 herd maintained under the high stocking rate (Coloraditas, Desiderio, and Guadalupe units) and 3 pastures, and 1 herd maintained under the moderate stocking rate (San Rafael, Loma, and Tequileras units)

Environmental predictors

We used canopy height, shrub density, grass spp. coverage, cacti spp. coverage, and bare ground coverage recorded from ground surveys in 2016 as environmental predictors in SDMs. 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, herbaceous (later classified by grass spp.), and bare ground 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).

Additionally, we used topographic relief (30-m^2 resolution) and Optimized Soil Adjusted Vegetation Index (OSAVI, a measure of LAI) produced from remotely sensed imagery collected during the same growing season as the ground surveys. We acquired one Landsat 8-OLI tile (< 6% cloud cover) that encompassed the study area (courtesy of U.S. Geological Survey) and processed this in ENVI 5.1 (NASA Landsat Program 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 image to the rectangular extent of the CGRDA pasture complex and then extracted by the study area mask in ESRI ArcGIS ArcMap 10.5. We then spatially subset the 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. 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, this vegetation index outperformed other, more common vegetation indices (e.g., Normalized Difference Vegetation Index [NDVI]) in

overall image classification accuracy and herbaceous coverage estimations (Fern et al. 2018).

Locations of water sources (e.g., livestock wells) within the study site and cattle stocking rates were provided by the East Foundation. To calculate water proximity, we gridded the spatial extent of the CGRDA into a fishnet (30-m^2 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 2016 (a total of 399 wells) were used in the analysis.

Quantifying grazing pressure

Several studies have cited the strong, predictable relationship between localized grazing pressure and proximity to water sources, especially in semi-arid rangelands (James et al. 1999; Landsberg and Crowley 2004; Locatelli et al. 2016; Ludwig et al. 2000). This spatially uneven use of the pasture by the livestock is even visible in satellite imagery as one study termed the zone of high livestock impact attenuating away from each water point (typically a livestock tank fed by a well) a "piosphere" (Andrew 1988). Piospheres are areas of high "hoof-action" and generally have higher accumulation of livestock feces, soil compaction, and defoliation (Andrew and Lange 1986; Graetz and Ludwig 1976). Due to the absence of natural water sources on the CGRDA, the known stocking rates of each pasture, and the well-documented relationship between localized grazing pressure and water sources (livestock tanks and wells) in semi-arid rangelands, we used water proximity to create a surrogate index for localized grazing pressure.

We used the distance to the nearest water source (livestock tank) previously calculated by the proximity analysis and 30-m^2 fishnet grid across the CGRDA. This ensured that resulting surface value estimates were the same spatial resolution as the other environmental rasters. We multiplied the distance value (m) of each fishnet pixel centroid by the density of grazing livestock (i.e., stocking rate) in each pasture using the raster math tool in ArcMap. A summary of all predictive layers included in the SDMs is presented in Table 1.

Bird occurrence data

Avian point counts consisted of 10 12-point transects (centrally located per pasture within the CGRDA). We used point count data collected on the CGRDA from

Table 1 Summary of predictor layers included in species distribution models (SDMs)

Predictor	Units
Canopy height	Meters
Shrub density	–
Grass spp. coverage	Proportion (%)
Cacti spp. coverage	Proportion (%)
Bare ground coverage	Proportion (%)
Topographic relief	Meters
OSAVI	–
Grazing pressure[a]	–

[a]Distance to the nearest water source (meters) multiplied by the cattle stocking rate of each pasture was used to represent grazingr pressure

April to June 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 occurrence records rather than abundance or density since the distributional modeling algorithms required presence/absence or presence only data. 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 disregarding the transect construct by subsampling the data by a 400-m cell size. This granted us a finer spatial resolution of the data set to thoroughly investigate the impacts of grazing pressure on grassland bird presence. We used only grassland-obligate species with an adequate number presence records within the CGRDA during the study period for distribution models: Northern Bobwhite, Eastern Meadowlark, and Cassin's Sparrow.

Data processing and analyses

We imported values for each predictor (canopy height, shrub density, bare ground coverage, grass spp. coverage, cacti spp. coverage, water proximity, and grazing pressure) into ArcMap 10.5 and used Kriging interpolation to minimize spatial sampling bias and create continuous surface layers of environmental predictor values. Kriging, or Gaussian process regression, 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 exported the resulting

Fern *et al. Ecological Processes*   (2020) 9:42

Page 5 of 11

GeoTIFFs and read these into the R statistical language as raster layers (Core Team 2013). We also read the GeoTIFFs representing the spatial values of elevation and topographic relief into R, and all layers were stacked to create the occurrence predictor rasters for the baseline SDMs.

We imported occurrence data for Northern Bobwhite, Eastern meadowlark, and Cassin's Sparrow into R and used the predictor raster stack to build SDMs using five different algorithms: BioClim (BC), generalized linear model (GLM), MaxEnt, boosted regression tree (BRT), and random forest (RF). We used the R package "dismo" to execute BioClim, generalized linear model, and MaxEnt (Hijmans et al. 2017). We used the R packages "gbm" and "randomForest" to execute boosted regression tree and random forest, respectively (Greenwell et al. 2019; Liaw and Wiener 2002). Table 2 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). These establish 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 for each species, we added the grazing pressure raster to the occurrence predictor raster stack and re-ran the models to assess any improvement or degradation in the predictive performance of each algorithm. Prior to building SDMs, we performed preliminary analyses for each species to ensure only predictors that added to the explanatory power of the models and did not add to the overall deviance which 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 species.

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

Table 2 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
BioClim (BC)	Presence only	This method uses a parallelepiped classifier to define species potential presence as the multi-dimensional environmental spaces bounded by the minimum and maximum values for all occurrences and gives a binary classification of suitable environment and unsuitable environment (Busby 1986, 1991).	Interpretations are straightforward, and the model is relatively simple to execute. More recently, this approach has proven useful in predicting biological invasions and distribution of widespread diseases (Robertson et al. 2004; Zhao et al. 2006).
Generalized linear (GLM)	Presence/absence	This is a generalization of the multiple regression model that uses the "link" function to accommodate non-linear relationships between the predictor and response variables. Using various transformations of the predictors (e.g., Logit, Poisson, and Gaussian) interactions between predictors can also be specified.	This approach is often ideal since occupancy modeling almost always involves multiple predictors, non-linear response functions, and response variables that are binary (Austin and Cunningham 198; Margules et al. 1987; 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 (Prasad et al. 2006; Gislason et al. 2006).
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 makes its application to ecological niche modeling ideal (Saatchi et al. 2008; Evangelista et al. 2009).
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 stage-wise 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 handing collinearity between predictor variables. BRT interpretations can be easily summarized to provide powerful ecological insight (Franklin 2010).

Fern *et al. Ecological Processes*   (2020) 9:42

Page 6 of 11

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 (e.g., any outcome above a 50% [0.50] likelihood indicates 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. This becomes especially meaningful when building SDMs for rare or endangered species that may have low prevalence across a given range or study area as the default threshold, usually 0.5, for many models (e.g., logistic regression-based GLM) may not be appropriate. In these cases, studies have suggested the use of binary species presence/absence maps as input which may be preferred for interpretation in building conservation plans, reservation networks, or sanctuaries as opposed to a continuous representation of probability of species presence (Fernández et al. 2006; Mladenoff and He 1999; Wilson et al. 2005). Although not prevalence independent, the AUC can be valuable in determining optimal threshold criteria. For example, Freeman and Moisen (2008) found that for SDMs projecting distributions of species with high prevalence (50%), default threshold criteria tended to converge. However, for species with low prevalence (e.g., 10%), the threshold where sensitivity + specificity is maximum offered the ideal probability threshold for species presence. In the R workspace output, this is typically read as "Max TPR + TNR" and can be exceedingly valuable for

accurately modeling distributions of rare or endangered species.

Results

We recorded a total of 1,565 occurrences for all three species within the CGRDA in the summer of 2016 (Northern Bobwhite = 996, Eastern Meadowlark = 179, Cassin's Sparrow = 390). Machine learning models (MaxEnt and RF) had the highest combinations of AUC and TSS for all species, with RF being the most consistent for each analysis (Table 3). In comparison of AUC values, the environmental envelope model (BC) and the GLM remained constant or only marginally improved with the addition of the grazing pressure raster. However, the TSS for these algorithms markedly improved with the addition of the grazing pressure raster for the Northern Bobwhite (ΔTSS = +0.93) and Eastern Meadowlark (ΔTSS = +0.08) SDMs (Table 3). The predictive power of both machine learning models and the BRT improved with the addition of the grazing pressure raster for all species, with the exception of MaxEnt and Eastern Meadowlark [Maxent: Northern Bobwhite [ΔAUC = +0.06], Cassin's Sparrow [ΔAUC = +0.02]; random forest: Northern Bobwhite [ΔAUC = +0.01], Eastern Meadowlark [ΔAUC = +0.05], Cassin's Sparrow [ΔAUC = +0.02]; random forest: Northern Bobwhite [ΔAUC = +0.03], Eastern Meadowlark [ΔAUC = +0.04], Cassin's Sparrow [ΔAUC = +0.03]. Random forest had the highest explanatory power (AUC) across all species but was, however, outperformed in sensitivity (TSS) by the other algorithms for all species for models including the grazing pressure raster (Table 3).

Northern Bobwhite distribution, the species of highest prevalence (*n* = 996), was best explained by random forest model inclusive of grazing pressure (AUC = 0.84; TSS = 0.48). However, the Bobwhite distribution was better explained by the addition of the grazing pressure raster by all algorithms as evidence in the measurable

Table 3 Results of species distribution model (SDM) performance for Bioclim, generalized linear model (GLM), MaxEnt, boosted regression tree (BRT), and random forest (RF) algorithms in predicting occurrence of Northern Bobwhite (NOBO), Eastern Meadowlark (EAME), and Cassin's Sparrow (CASP) on East Foundation's Coloraditas Grazing Research and Demonstration Area (CGRDA) in the summer of 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 localized grazing pressure (denoted by "+").

		BioClim		GLM		MaxEnt		BRT		RF	
		AUC	TSS	AUC	TSS	AUC	TSS	AUC	TSS	AUC	TSS
NOBO		0.54	0.032	0.64	0.034	0.61	0.056	0.59	0.049	0.81	0.42
	+	0.58	0.96	0.67	0.24	0.67	0.18	0.60	0.57	0.84	0.48
EAME		0.81	0.41	0.78	0.62	0.79	0.61	0.84	0.46	0.91	0.62
	+	0.81	0.49	0.78	0.80	0.78	0.75	0.89	0.80	0.95	0.67
CASP		0.58	0.13	0.44	0.18	0.69	0.24	0.62	0.21	0.78	0.22
	+	0.62	0.15	0.44	0.18	0.71	0.29	0.64	0.67	0.81	0.23

increase in AUC and TSS in each model (ΔAUC = + 0.01–0.06, ΔTSS = + 0.04–0.93; Table 3).

Eastern Meadowlark distribution, the species of the lowest prevalence (n = 179), was also best explained by the random forest model inclusive of grazing pressure (AUC = 0.95; TSS = 0.67). The SDM explanatory power for this species' distribution was not improved with the addition of grazing pressure using the BioClim, GLM, and MaxEnt algorithms. Cassin's Sparrow distribution, the species of moderate prevalence (n = 390), was also best explained by the random forest model inclusive of grazing pressure (AUC = 0.81; TSS = 0.23). However, the SDM explanatory power for this species' distribution was not improved with the addition of grazing pressure using the GLM algorithm. Additionally, other algorithms (BRT and MaxEnt) produced higher TSS values (TSS = 0.67 and 0.29, respectively).

Discussion

Our novel approach to spatially quantify localized grazing pressure improved the prediction accuracy and sensitivity of SDMs projecting the distribution of Northern Bobwhite, Eastern Meadowlark, and Cassin's Sparrow. Of the three algorithms used, random forest performed best for explaining presence regardless of species prevalence and should be preferred by rangeland managers seeking to promote sustainable livestock grazing while balancing the needs of sensitive wildlife populations. Random forest models operate on a machine-learning, decision tree mechanism. Thus, the superior performance of RF in this study implies that it is a valuable approach to limited, binary data (e.g., presence/absence). It is important to note the varying model performance with relation to species prevalence. For example, SDMS built to project distributions of Northern Bobwhite, the species of the highest prevalence in this study varied widely in predictive performance (AUC) and sensitivity (TSS) across algorithms. Rangeland managers should consider both metrics (AUC and TSS) when assessing model performance since both provide valuable insight into the over utility of the model (i.e., AUC describing explanatory power and TSS describing model stability or sensitivity to the predictors). Although both AUC and TSS are theoretically prevalence independent, for species like Northern Bobwhite that are often locally abundant where they are present, machine-learning models that can accommodate non-linear relationships (e.g., random forest) should be preferred in modeling distributions. In an ecological context, the improvement in model sensitivity and explanatory power seen with the addition of grazing pressure to Northern Bobwhite SDMs should be considered meaningful by rangeland ecologists. The direct impacts of livestock grazing (e.g., changes in vegetative structure and composition) on the distribution of Northern Bobwhite is well recognized (Baker and Guthery 1990; Coppedge et al. 2008; Flanders et al. 2006; Lusk et al. 2002). However, with the inclusion of grazing pressure as an indirect variable and the subsequent increase in explanatory power across all algorithms (ΔAUC = + 0.01–0.06), our findings suggest this species' distribution is also indirectly affected by livestock grazing activities. Thus, future investigations into the Northern Bobwhite distribution or populations should consider the presence and localized intensity of livestock grazing.

The addition of grazing pressure as a variable also increased the explanatory power and sensitivity of some SDMs built to project distributions of Cassin's Sparrow, the species of moderate prevalence in this study (BioClim, MaxEnt, BRT, RF). However, any improvements in the model performance were marginal (ΔAUC = + 0.0–0.4). Our findings suggest indirect effects of livestock grazing on Cassin's Sparrow presence, though marginally detectable, were negligible. Rangeland managers should consider the unique ecological circumstances of each rangeland and livestock grazing system when investigating Cassin's Sparrow distribution or presence. Although both machine-learning models (MaxEnt and random forest) and boosted regression tree performed relatively well, compared to the envelope (BioClim) and logistic algorithms (generalized linear model), the BRT produced the highest model sensitivity. This is likely due to the innate accommodation of missing and limited data in this algorithm, which makes it ideal for species of lower (or unknown) prevalence. In these cases, the boosted regression tree provides a superior, yet conservative SDM for rangeland ecologists seeking to project distributions of species with low to moderate or unknown prevalence.

Distributions of Eastern Meadowlark, the species of the lowest prevalence in this study, were better explained by the addition of grazing pressure only in the boosted regression tree and random forest SDMs. Although previous studies have suggested a neutral effect of livestock grazing activity on the presence of Eastern Meadowlark, this species has also been known to alter behavior and be particularly susceptible to brood parasitism (usually by Brown-headed cowbird *Molothrus ater*) in heavily grazed pastures (Baker and Guthery 1990; Coppedge et al. 2008). Further, Roseberry and Klimstra (1970) found substantial differences in Eastern Meadowlark nest densities between lightly grazed and heavily grazed pastures of similar vegetation composition and area. While direct impacts of livestock grazing (e.g., changes in vegetative structure) may not be as evident in the distributions of this species as they are in others (e.g., Northern Bobwhite), our findings suggest some indirect influence of livestock grazing activity on Eastern Meadowlark presence. The random forest algorithm, in

the accommodation of missing data and low presence values, produced the SDM with the highest explanatory power for this species, and it should be preferred for other species of low prevalence.

In a broader context, our investigation of grazing pressure influence on bird distributions could be expanded to explore the effect of other land uses (e.g., urbanization and cultivation) on distributions. For instance, the effect of native prairie versus cultivated cropland on the distribution of birds of the same species or the more complex interaction between disturbance and brood parasites (e.g., Brown-headed cowbird *Molothrus ater*) and that effect on the distribution of susceptible bird species (Brittingham and Temple 1983). This becomes especially relevant as climate change accelerates the impacts of land-use changes across the landscape (Pielke 2005).

BioClim
This algorithm is traditionally used as an environmental envelope method to model large scale distributions and invasions (Hijmans et al. 2001, 2005). However, recent improvements in the algorithm (in the R package "Dismo" [Hijmans et al. 2017]) have allowed analyses of single species occurrences at finer resolutions. The binary output also makes it especially well-suited for species with low prevalence. For example, it performed best (AUC = 0.81) with the Eastern Meadowlark, the species of the lowest prevalence in this study. For this species, this model did not improve with the addition of grazing pressure as a predictor. Since other models showed improvement with the addition of grazing pressure (BRT and RF), this may suggest some disadvantage to the linearity of this algorithm. BioClim also had the poorest predictive performance (AUC = 0.54; 0.58, with and without grazing pressure, respectively) for Northern Bobwhite. This species had the highest prevalence in the study and, thus, may suggest a saturation limitation for this algorithm as large sample sizes have been recognized to de-stabilize similar models (Mateo et al. 2010).

GLM (binomial)
The SDMs built using this logistic regression-based algorithm, generally, performed poorly, especially for Cassin's Sparrow (AUC = 0.44). Additionally, GLM SDMs for Eastern Meadowlark and Cassin's Sparrow did not improve with the addition of grazing pressure despite the improvement seen in other models. Although this algorithm can theoretically accommodate non-linear relationships between predictor and response variables, it has been recognized to over-fit distribution models producing biased or inaccurate results (Austin and Cunningham 1981; Elith and Graham 2009).

MaxEnt
SDMs built using this machine-learning algorithm projecting Northern Bobwhite, and Cassin's Sparrow distributions improved with the addition of grazing pressure as a predictor. However, predictive power of the Eastern Meadowlark SDM decreased with the addition of grazing pressure (AUC = 0.79, 0.78; respectively) while the TSS remained high (0.61, 0.75, respectively). Although not a rare or endangered species, this was the species of the lowest prevalence in the study and supports the concept suggested by Freeman and Moisen (2008) that default probability thresholds may not be appropriate at low prevalence, and that the intersection where Sensitivity + Specificity is maximum could serve as a more ideal probability threshold for species presence. We did not perform this analysis here but is an area of interest for future research in improving SDMs.

BRT
The BRT performed best with Eastern Meadowlark SDMs (AUC = 0.89), and all species' models improved with the addition of grazing pressure as a predictor. This algorithm has the unique advantage to accommodate collinearity among predictors and fit complex nonlinear relationships between response and predictor variables (Elith et al. 2008; Franklin 2010). Among the SDMs projecting Cassin's Sparrow distribution, the BRT had the highest model sensitivity (TSS = 0.67). The BRT requires two user-input parameters: learning rate (lr), which determines the contribution of each decision tree to the overall model, and tree complexity (tc), which controls whether interactions are fitted (Elith et al. 2008). Ideally, parameters should be optimized based on sample size, number of predictors, intended use of the model, etc. to avoid overfitting the model. However, for the purposes of this study, we maintained consistent parameters to directly compare model performance ($lr = 0.001$, $tc = 6$). This may have contributed to the poor predictive performance of the BRT in projecting Northern Bobwhite distribution relative to the other two species.

RF
This regression-based machine-learning algorithm performed best for Eastern Meadowlark SDMs (AUC = 0.95) and produced the most powerful SDMs for all species. All models built using this algorithm improved with the addition of grazing pressure as a predictor, and model sensitivity was relatively consistent compared to the output of the other SDMs. Whereas the BRT requires the user to alter input parameters to ensure the model is not over fitted, RF has the advantage of a built-in "safe-guard" against overfitting in that each decision tree which uses a random bootstrap aggregation to subsample the given data (Breiman 2001; Prasad et al.

2006). RF is growing in popularity among ecologists for SDM and shows great promise for advanced SDM applications since it makes no assumptions on data distributions.

Conclusions

Our findings suggest that model selection for SDM should include consideration of species prevalence and machine-learning algorithms should be preferred when the target species is of low or unknown prevalence. For example, rangeland ecologists building SDMs for a species that is either rare across its range or of unknown abundance are able to select or alter the probability threshold of species presence in machine-learning algorithms. This is especially valuable since SDMs build based on the default probability threshold (0.5) used for rare or endangered species could lead to misinformed conservation plans and refuge networks. This new approach in spatially quantifying and including livestock grazing pressure as an indirect variable in SDMs has broad implications in rangeland ecology since it addresses a weakness in the current SDM framework—the exclusion of biotic and indirect relationships. With this, we can better estimate the effects of varying grazing regimes on grassland bird populations and more accurately predict the distribution of species of interest

Further, our results imply livestock grazing has indirect influence on grassland bird species' distributions and should be included in SDMs as an indirect variable in addition to direct, associated vegetative changes. This is especially important for ground-dwelling species (e.g., Northern Bobwhite). For instance, more advanced boosting or machine-learning algorithms (e.g., boosted regression tree and random forest) that can accommodate limited data, complex and non-linear relationships, and collinearity among predictors could inform a rangeland ecologist if the redistribution, or absence of breeding quail on a property, is more heavily influenced by the absence of rainfall during drought conditions (an indirect effect) or the resulting senescence of vegetation (a direct effect of drought). Algorithms that can tease apart these effects can help inform effective, science-based management.

Authors' contributions
RF conceptualized the study design, analyzed and interpreted data, and wrote the manuscript text. MM made major contributions to the study design implementation and inferences from the analyses. WG contributed to the design of the statistical analyses. HW contributed to the interpretation of model comparison and selection as well as the writing of the manuscript text. TC provided vital guidance in manipulating bird observation data and generating vegetation survey data used in this study. The authors read and approved the final manuscript.

Availability of data and materials
Bird observation, vegetation, and grazing regime data used in this study are not publicly available due to their origination on a private property. Data are available from the corresponding author on reasonable request.
Imagery data used in this study are publicly available as follows: 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. Climate data used in this study are publicly available as follows: PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu. Hunting lease data used in this study are publicly available as follows: Texas Parks and Wildlife Department [TPWD]. [Hunt Texas] calculated from "All Lease Listings" on 16 April 2017 (https://www2.tpwd.state.tx.us/huntwild/hunt/planning/hunt_lease/listlease.php).

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Wildlife and Fisheries Sciences, Texas A&M University, WFES Building, 2258 TAMU, 534 John Kimbrough Blvd, College Station, TX 77843-2258, USA. [2]East Foundation, San Antonio, TX 78216, USA.

References
Andrew MH (1988) Grazing impact in relation to livestock watering points. Trends Ecol Evol 3:336–339. https://doi.org/10.1016/0169-5347(88)90090-0

Andrew MH, Lange RT (1986) Development of a new piosphere in arid chenopod shrubland grazed by sheep. 1. Changes to the soil surface. Austral Ecol 11:395–409. https://doi.org/10.1111/j.1442-9993.1986.tb01409.x

Askins RA, Chávez-Ramírez F, Dale BC, Haas CA, Herkert JR, Knopf FL, Vickery PD (2007) Conservation of grassland birds in North America: understanding ecological processes in different regions:"Report of the AOU Committee on Conservation". Ornithol Monogr 64:1–46

Atauchi PJ, Peterson AT, Flanagan J (2018) Species distribution models for Peruvian plantcutter improve with consideration of biotic interactions. J Avian Biol 49:jav-01617

Austin MP, Cunningham RB (1981) Observational analysis of environmental gradients. Proc Ecol Soc Aust 11:109–119

Austin MP, Van Niel KP (2011) Improving species distribution models for climate change studies: variable selection and scale. J Biogeogr 38:1–8

Baker DL, Guthery FS (1990) Effects of continuous grazing on habitat and density of ground foraging birds in South Texas. J Range Manag 43:2–5. https://doi.org/10.2307/3899109

Bonham CD (2013) Measurements for terrestrial vegetation, 2nd edn. Wiley, Chichester

Breiman L (2001) Random forests. Mach Learn 45:5–32

Brennan LA, Kuvlesky WP Jr (2005) North American grassland birds: an unfolding conservation crisis? J Wildl Manage 69:1–13. https://doi.org/10.2193/0022-541X(2005)069<0001:NAGBAU>2.0

Brittingham MC, Temple SA (1983) Have cowbirds caused forest songbirds to decline? BioScience 33:31–35

Busby J (1991) BIOCLIM - a bioclimate analysis and prediction system. In: Margules CR, Austin MP (eds) Nature conservation: cost effective biological surveys and data Analysis. CSIRO, East Melbourne, pp 64–68

Busby JR (1986) A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. Aust J Ecol 11:1–7

Canfield RH (1941) Application of the line interception method in sampling range vegetation. J For 39:388–394

Coppedge BR, Fuhlendorf SD, Harrell WC, Engle DM (2008) Avian community response to vegetation and structural features in grasslands managed with fire and grazing. Biol Conserv 141:1196–1203

Core Team R (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna http://www.R-project.org/

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 Tech Ref 1734, pp 1–171

Derner JD, Lauenroth WK, Stapp P, Augustine DJ (2009) Livestock as ecosystem engineers for grassland bird habitat in the western Great Plains of North America. Rangel Ecol Manag 62:111–118

Dodd EP (2009) An expense and economic impact analysis of hunting operations in South Texas. In: Masters Abstracts International 48 No. 01

Elith J, Graham CH (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32:66–77

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677–697

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802–813

Evangelista PH, Stohlgren TJ, Morisette JT, 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 RR, Foxley EA, Bruno A, Morrison ML (2018) Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland. Ecol Indic 94:16–21

Fernández N, Delibes M, Palomares F (2006) Landscape evaluation in conservation: molecular sampling and habitat modeling for the Iberian lynx. Ecol Appl 16:1037–1049

Flanders AA, Kuvlesky WP Jr, Ruthven DC III, Zaiglin RE, Bingham RL, Fulbright TE, Hernández F, Brennan LA (2006) Effects of invasive exotic grasses on south Texas rangeland breeding birds. Auk 123:171–182

Franklin J (2010) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge

Freeman EA, Moisen GG (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol Model 217:48–58

Fuhlendorf SD, Harrell WC, Engle DM, Hamilton RG, Davis CA, Leslie DM (2006) Should heterogeneity be the basis for conservation? Grassland bird response to fire and grazing. Ecol Appl 16:1706–1716

Fulbright TE, Diamond DD, Rappole J (1990) The coastal sand plain of southern Texas. Rangelands 12:337–340

Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. Pattern Recognit Lett 27:294–300

Graetz RD, Ludwig JA (1976) A method for the analysis of piosphere data applicable to range assessment. Rangel J 1:126–136

Greenwell B, Boehmke B, Cunningham J, GBM Developers (2019) gbm: generalized boosting regression models. R package version 2.1.5. https://CRAN.R-project.org/package=gbm.

Higgins KF, Oldemeyer JL, Jenkins KJ, Clambey GK, Harlow RF (1996) Vegetation sampling and measurement. Res Manag Tech Wildl Habitats 5:567–591

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978

Hijmans RJ, Guarino L, Cruz M, Rojas E (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. Plant Genet Resour News l:15–19

Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) dismo: Species distribution modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo

James CD, Landsberg J, Morton SR (1999) Provision of watering points in the Australian arid zone: a review of effects on biota. J Arid Environ 41:87–121

Jansen R, Little RM, Crowe TM (1999) Implications of grazing and burning of grasslands on the sustainable use of francolins (*Francolinus* spp.) and on overall bird conservation in the highlands of Mpumalanga province, South Africa. Biodivers Conserv 8:587–602

Knopf FL (1994) Avian assemblages on altered grasslands. Stud Avian Biol 15:247–257

Landsberg J, Crowley G (2004) Monitoring rangeland biodiversity: plants as indicators. Austral Ecol 29:59–77

Leach K, Montgomery WI, Reid N (2016) Modelling the influence of biotic factors on species distribution patterns. Ecol Model 337:96–106

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2:18–22

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. https://doi.org/10.1111/jbi.12058

Locatelli AJ, Mathewson HA, Morrison ML (2016) Grazing impact on brood parasitism in the Black-Capped Vireo. Rangel Ecol Manag 69:68–75

Ludwig JA, Bastin GN, Eager RW, Karfs R, Ketner P, Pearce G (2000) Monitoring Australian rangeland sites using landscape function indicators and ground- and remote-based techniques. In: Monitoring Ecological Condition in the Western United States. Springer, Heidelberg, pp 167–178

Lusk JJ, Guthery FS, George RR, Peterson MJ, DeMaso SJ (2002) Relative abundance of bobwhites in relation to weather and land use. J Wildl Manage 66:1040–1051

Mahmoudabadi H, Briggs G (2016) Directional kriging implementation for gridded data interpolation and comparative study with common methods. AGU Fall Meeting Abstracts.

Margules CR, Nicholls AO, Austin MP (1987) Diversity of *Eucalyptus* species predicted by a multi-variable environmental gradient. Oecologia 71:229–232

Mateo RG, Felicísimo ÁM, Muñoz J (2010) Effects of the number of presences on reliability and stability of MARS species distribution models: the importance of regional niche variation and ecological heterogeneity. J Veg Sci 21:908–922. https://doi.org/10.1111/j.16541103.2010.01198.x

Miller J (2010) Species distribution modeling. Geogr Compass 4:490–509. https://doi.org/10.1111/j.1749-8198.2010.00351.x

Mladenoff DJ, He HS (1999) Design, behavior and application of LANDIS, an object-oriented model of forest landscape disturbance and succession. In: Mladenoff DJ, Baker WL (eds), Spatial Modeling of Forest Landscape Change: Approaches and Applications. Cambridge University Press, Wisconsin, p 125–162

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)

Nocera JJ, Koslowsky HM (2011) Population trends of grassland birds in North America are linked to the prevalence of an agricultural epizootic in Europe. Proc Natl Acad Sci 108:5122–5126

Phillips S, Elith J (2011) Logistic methods for resource selection functions and presence-only species distribution models. In: Twenty-Fifth AAAI Conference on Artificial Intelligence, pp 1384–1389

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Phillips SJ, Dudik M, Elith J, Graham CH, 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. https://doi.org/10.1890/07-2153.1

Pielke RA (2005) Land use and climate change. Science 310:1625–1626

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199

PRISM Climate Group. Oregon State University. 2018. http://prism.oregonstate.edu.

Renner IW, Warton DI (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. Biometrics 69:274–281

Robertson MP, Villet MH, Palmer AR (2004) A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. Divers Distrib 10:461–474

Rondeaux G, Steven M, Baret F (1996) Optimization of soil-adjusted vegetation indices. Remote Sens Environ 55:95–107

Roseberry JL, Klimstra WD (1970) The nesting ecology and reproductive performance of the Eastern Meadowlark. Wilson Bull 82:243–267

Saatchi S, Buermann W, Steege HT, Mori S, Smith TB (2008) Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. Remote Sens Environ 112:2000–2017

Snelgrove A, Dube A, Skow K, Engeling A (2013) East Wildlife Foundation Atlas. Texas A&M Institute of Renewable Natural Resources, College Station

Texas Parks and Wildlife Department [TPWD] (2016) South Texas Wildlife Management. http://tpwd.texas.gov/landwater/habitats/southtx_plain . Accessed 16 Apr 2016

Texas Parks and Wildlife Department [TPWD] (2017) [Hunt Texas] calculated from "All Lease Listings". https://www2.tpwd.state.tx.us/huntwild/hunt/planning/hunt_lease/listlease.php

Van Beers WC, Kleijnen JP (2004) Kriging interpolation in simulation: a survey. In: Proceedings of the December 36th Conference on Winter Simulation. IEEE 1: 113–121.

Ward G, Hastie T, Barry S, Elith J, Leathwick JR (2009) Presence-only data and the EM algorithm. Biometrics 65:554–563

Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biol Conserv 122:99–112

Zhao C, Nan Z, Cheng G, Zhang J, Feng Z (2006) GIS-assisted modelling of the spatial distribution of Qinghai spruce (*Picea crassifolia*) in the Qilian Mountains, northwestern China based on biophysical parameters. Ecol Model 191:487–500

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.