Integrating anthropogenic factors into regional-scale species distribution models—A novel application in the imperiled sagebrush biome

Juan M. Requena-Mullor | Kaitlin C. Maguire | Douglas J. Shinneman | Timothy Trevor Caughlin

Abstract

Species distribution models (SDMs) that rely on regional-scale environmental variables will play a key role in forecasting species occurrence in the face of climate change. However, in the Anthropocene, a number of local-scale anthropogenic variables, including wildfire history, land-use change, invasive species, and ecological restoration practices can override regional-scale variables to drive patterns of species distribution. Incorporating these human-induced factors into SDMs remains a major research challenge, in part because spatial variability in these factors occurs at fine scales, rendering prediction over regional extents problematic. Here, we used big sagebrush (Artemisia tridentata Nutt.) as a model species to explore whether including human-induced factors improves the fit of the SDM. We applied a Bayesian hurdle spatial approach using 21,753 data points of field-sampled vegetation obtained from the LANDFIRE program to model sagebrush occurrence and cover by incorporating fire history metrics and restoration treatments from 1980 to 2015 throughout the Great Basin of North America. Models including fire attributes and restoration treatments performed better than those including only climate and topographic variables. Number of fires and fire occurrence had the strongest relative effects on big sagebrush occurrence and cover, respectively. The models predicted that the probability of big sagebrush occurrence decreases by 1.2% (95% CI: −6.9%, 0.6%) when one fire occurs and cover decreases by 44.7% (95% CI: −47.9%, −41.3%) if at least one fire occurred over the 36 year period of record. Restoration practices increased the probability of big sagebrush occurrence but had minimal effect on cover. Our results demonstrate the potential value of including disturbance and land management along with climate in models to predict species distributions. As an increasing number of datasets representing land-use history become available, we anticipate that our modeling framework will have broad relevance across a range of biomes and species.

Keywords

Anthropocene, Artemisia tridentata, Bromus tectorum, ecological restoration, global change, Great Basin, invasive species, large-scale, sagebrush steppe, wildfire
INTRODUCTION

Anthropogenic environmental change is occurring at multiple scales, including regional and global, while management actions to conserve and restore ecosystem function are typically limited to site- and landscape scales. This mismatch presents a major challenge for biodiversity conservation, as logistical considerations limit the scale of most management actions, including restoration of degraded ecosystems and control of wildfire. Even where interventions over large spatial extents are possible, there is considerable uncertainty in how management actions developed at site-level scales will perform across heterogeneous landscapes (Holl, Crone, & Schultz, 2003; King & Keeland, 1999). Spatial models could provide a solution to these challenges by integrating data across scales to provide quantitative scenarios for ecological impacts of human activity. Nevertheless, models that simultaneously include regional-scale climatic and topographic variables as well as finer-scale impacts of human activity remain scarce (see for example Greve, Lykke, Blach-Overgaard, & Svenning, 2011; Lippitt et al., 2008; Veran et al., 2016).

Species distribution models (SDMs) provide an example of a modeling tool designed to explore how environmental conditions influence the spatial pattern of species occurrence over large areas. SDMs suit a range of decision-making contexts and are a cornerstone of conservation biology (see Franklin, 2013 for a review). Most SDMs rely on regional-scale environmental variables derived from climate, topography, and land cover (Miller, 2010). These covariates vary predictably along continental-scale environmental gradients and can be measured reliably at a variety of scales, facilitating spatial comparisons. However, in the Anthropocene, species’ distributions are increasingly overidden by human influence, including altered wildfire regimes, land-use change, and invasive species (Crimmins, Dobrowski, Mynsberge, & Safford, 2014; Hellmann, Werner, & Oldeland, 2016; Jetz, Wilcove, & Dobson, 2007; Lewis & Maslin, 2015). Incorporating these human-induced factors into SDMs at a regional scale remains a major challenge. First, these factors are outcomes of coupled natural-human systems and often vary unpredictably across scales, in part due to heterogeneity in land manager decision-making (Mangiacci et al., 2013; Torres-Romero & Olalla-Tárraga, 2015). Second, while standardized measurements of climate and topography are available at global scales, datasets on human-induced land change tend to be idiosyncratic and sparsely available, rendering prediction of human impacts over regional extents problematic (Díez & Pulliam, 2007; Kelly, Leach, Cameron, Maggs, & Reid, 2014).

Fire is a disturbance affecting ecosystems over a range of scales, from individual plants to multiple watersheds, and contemporary fire regimes are increasingly altered by anthropogenic environmental change (Dennison, Brewer, Arnold, & Moritz, 2014). Fire can influence the distribution of individual species over time and space (Bond & Keeley, 2005). However, the few studies that have directly included fire attributes in SDMs have reached different conclusions regarding its utility. Tucker, Rebelo, and Manne (2012) included fire attributes into SDMs of species in the South African fynbos community and found that the inclusion of fire increased model performance. Sherrill and Romme (2012) modeled the distribution of cheatgrass (Bromus tectorum L.) in the northern Colorado Plateau and found that including fire severity as a covariate increased model performance, but the effect varied with the spatial scale of analysis. In contrast, Crimmins et al., (2014) found that the inclusion of fire did not increase predictive ability of SDMs of vascular plants throughout California. Incorporating fire as a covariate within a SDM will require careful consideration of the specific systems being modeled, key predictor attributes, response variable types, and the potential role of confounding variables. For instance, the different conclusions regarding the utility of fire among the previous SDM studies (Crimmins et al., 2014; Sherrill & Romme, 2012; Tucker et al., 2012) may be due to factors unique to plant-fire relationships within each study region, whether species responses were modeled based on abundances or presence/absence, and the precise fire attributes used in the models (e.g., occurrence, frequency, size, severity). Predicting which species have distributions that will be altered by fire is of paramount importance, especially because climate change may potentially increase the number, size, and intensity of wildfires (Holden et al., 2018; Schoennagel et al., 2017).

As anthropogenic disturbances increase in scale, so have calls for increasing the scale of ecological restoration (Aronson & Alexander, 2013). Broad-scale plans to restore ecosystem function to extensive amounts of degraded land include the Aichi Convention on Biodiversity (global restoration of 15% of degraded ecosystems; Convention on Biological Diversity, 2010) and the Bonn Challenge (restoration of 150 million hectares; http://www.bonnchallenge.org/). However, achieving restoration at landscape to regional scales will require spatial planning to allocate limited resources for effective results (Barbosa & Asner, 2017; Neeson et al., 2015), including the need to consider climate variation in restoration plans (Harris, Hobbs, Higgs, & Aronson, 2006; Jackson & Hobbs, 2009). Most studies for this purpose have taken a prospective approach to forecast restoration outcomes based on climatic variation in plant performance (Amburagey et al., 2018; Harris et al., 2006; Richardson & Chaney, 2018). For example, Doherty, Butterfield, and Wood (2017) used climate-based SDMs to predict relative performance of native plant materials for restoration projects across the western U.S. In contrast, retrospective analyses that quantify the success (or failure) of regional-scale restoration projects are rare (Shoo et al., 2017), even though high variability in restoration outcomes over space and time is the norm (Brudvig et al., 2017; Jones et al., 2018). Indeed, there is a high likelihood that local-scale anthropogenic factors will override the importance of regional-scale climate variation for success of restoration projects. Testing the relative importance of these factors at different scales will require incorporating climate variation into retrospective analyses of restoration treatment effectiveness. However, to our knowledge, no SDMs have united restoration treatment history and regional-scale abiotic variation.
Sagebrush (Artemisia spp.)-dominated landscapes of western North America present ideal ecosystems for testing the inclusion of anthropogenic variables into SDMs. Sagebrush taxa and ecosystems are widespread, historically extending across ~1 million km$^2$ of arid and semiarid environments in the western U.S. (Beetle, 1960), but are threatened by a variety of anthropogenic influences, including clearing for agriculture, altered fire regimes, and expansion of exotic species (Miller et al., 2011). Disturbance by wildfire has become the major threat to sagebrush steppe and shrublands in the western half of the sagebrush biome, in large part due to the influence of invasive grasses and forbs, especially cheatgrass, that increase fine fuels and promote frequent fire (Balch, Bradley, D’Antonia, & Gómez-Dans, 2013; Figure 2). Big sagebrush does not resprout and has limited seed dispersal capacity, so it does not recover quickly after fire (Shinneman & McIlroy, 2016) and postfire recovery of stands is via sexual reproduction (Shriver et al., 2018). While native sagebrush habitats historically supported mean fire return intervals that ranged roughly half of the sagebrush biome in terms of the total area historically covered by sagebrush-dominated ecosystems. Climate generally varies from semiarid to arid in the region, with warm to hot summers and cold winters. The mean annual temperature averaged across the study area is 9.10°C (±3.70 SD) and the cumulative annual precipitation is 370.21 mm (±248.83 SD) (Daly et al., 2008); however, temperature and precipitation generally decrease and increase, respectively, with higher elevation and latitude. The landscape is dominated by basin and range topography, with abrupt elevation changes from valleys to mountains (elevation range is 4,416 m) that support similarly abrupt changes in species composition. From lowest to highest elevations, the natural communities of the Great Basin generally transition from salt desert shrublands, to sagebrush steppe or shrublands, to a mixture of sagebrush and pinyon–juniper woodlands, to isolated, high-mountain forests and alpine communities. Sagebrush communities were historically extensive within low- to mid-elevation landscapes of the Great Basin. However, they have lost considerable area over the past several decades due to altered fire regimes, primarily driven by expansion of flammable, invasive annual grasses and forbs (as described above).

### 2 | MATERIAL AND METHODS

#### 2.1 | Study area

The study area is the Great Basin of North America (Figure 1), a ‘cold-desert’ ecoregion, defined here by hydrographic boundaries (U.S. Geologic Map, 1992) and expanded with a 100 km buffer zone to cover approximately 400,000 km$^2$. The Great Basin constitutes roughly half of the sagebrush biome in terms of the total area historically covered by sagebrush-dominated ecosystems. Climate generally varies from semiarid to arid in the region, with warm to hot summers and cold winters. The mean annual temperature averaged across the study area is 9.10°C (±3.70 SD) and the cumulative annual precipitation is 370.21 mm (±248.83 SD) (Daly et al., 2008); however, temperature and precipitation generally decrease and increase, respectively, with higher elevation and latitude. The landscape is dominated by basin and range topography, with abrupt elevation changes from valleys to mountains (elevation range is 4,416 m) that support similarly abrupt changes in species composition. From lowest to highest elevations, the natural communities of the Great Basin generally transition from salt desert shrublands, to sagebrush steppe or shrublands, to a mixture of sagebrush and pinyon–juniper woodlands, to isolated, high-mountain forests and alpine communities. Sagebrush communities were historically extensive within low- to mid-elevation landscapes of the Great Basin. However, they have lost considerable area over the past several decades due to altered fire regimes, primarily driven by expansion of flammable, invasive annual grasses and forbs (as described above).

#### 2.2 | Species model and occurrence data

Occurrence (presence and absence) and cover data for *A. tridentata* (big sagebrush) were obtained from the LANDFIRE Reference Database (https://www.landfire.gov/lfrdb.php), that contains plot-based species composition and percentage cover data derived from field-identification for the period 1980 through 2015. Big sagebrush was considered as ‘absent’ when the percentage cover was zero. While the three primary subspecies of *A. tridentata* likely have unique niches (e.g., Still & Richardson, 2015), we modeled at the species level because field identification of sagebrush subspecies is a suspect due to both morphological and genetic ambiguities among putative subspecific populations (McArthur, Welch, & Sanderson, 1988; Richardson, Page, Bajgian, Sanderson, & Udall, 2012). To remove potential noise, we filtered the data by: (a) removing erroneous records which were outside of the known distribution range of the species; (b) selecting only records with at least five species identified...
in a sampled location to ensure that the absence records were as robust as possible; (c) removing records located within individual fire boundaries that were collected before a fire burned (to reflect the influence of fire, see Fire data subsection); and similarly (d) discarding records within treatment area polygons that were collected before a treatment occurred. After filtering, 13,002 presence and 8,751 absence points for *A. tridentata* remained, resulting in 21,753 total data points.

2.3 | Fire data

Fire attributes were extracted from a historical fire dataset compiled from various federal, state, and local sources (Welty, Pilliod, & Arkle, 2017). Given uncertainty concerning increasingly older fire records, we restricted our analysis to a recent 36 year period (1980–2015), during which fire record omissions are relatively few and fire perimeters are more accurate. Two fire attributes were computed: fire occurrence (FO, binary variable with 0 and 1 representing fire absence and presence, respectively) and number of fires (NOF, number of fires between 1980 and 2015; ranging from 0 to 3 fires). Because fire typically kills sagebrush and ostensibly influences its postfire distribution (Figure 2), we removed LANDFIRE data points that fell within fire perimeters but were surveyed before the date of the last fire for a specific location (i.e., to remove potentially false extant occurrences), but retained points within fire perimeters that were established after the last fire date (i.e., to include the potential influence of fire on extant sagebrush distribution). Such filtering resulted in 12,293 presence and 8,268 absence points for *A. tridentata* with no fire, and 709 presence and 483 absence points with fire.

2.4 | Land Treatment Digital Library

Treatment data were obtained from the Land Treatment Digital Library (LTDL; https://ltdl.wr.usgs.gov/). The LTDL catalogs legacy land treatment information on Bureau of Land Management lands in the western U.S. and contains more than 42,000 spatially explicit records of individual treatments (Pilliod & Welty, 2013). We considered treatment polygons that had a restoration focus and would potentially benefit sagebrush recovery, including sagebrush plantings and seedings, postfire soil stabilization, livestock closure, and invasive species control (see Table S1). To create the treatment factor, the LANDFIRE records within the treated polygons were labeled as either ‘Treatment’ or ‘No Treatment’. We grouped all restoration treatments into one level because there were not an adequate number of LANDFIRE points within each treatment type to robustly detect their potentially unique effects. This process yielded 1,250 presence and 536 absence points for *A. tridentata* labeled as—Treatment, and 11,752 presence and 8,215 absence points labeled as—No treatment.

2.5 | Climate and topographic covariates

We selected two 30 year average (1981–2010) climate covariates: mean annual temperature (in °C; TEMP) and cumulative annual
precipitation (in mm; PREC); and one topographic variate (in degrees; SLOPE). Such covariates have well-established effects on sagebrush occurrence (Kleinhesselink & Adler, 2018; Tredennick et al., 2016) and a Spearman rank correlation <0.7 (Dormann et al., 2013). Climate variables were derived from monthly data available from the Parameter-elevation Regression on Independent Slopes Model (PRISM; Daly et al., 2008) at 800 m resolution. It is important to note here that our goal was not to create a highly predictive SDM based on climate variables, but rather to test how the inclusion of anthropogenic factors could improve climate-based SDM models. Accordingly, we selected ‘primary’ climate covariates instead of more nuanced ones, such as seasonal or time-lag covariates.

SLOPE (steepness in degrees) was computed in Google Earth Engine (Gorelick et al., 2017) using the digital elevation data at a resolution of 1 arc-second (approximately 30 m resolution) from the Shuttle Radar Topography Mission (Farr et al., 2007). Slope affects sagebrush occurrence by determining the local conditions of aridity, that is, soil depth and texture, solar radiation, and wind exposure (Burke, Reiners, & Olson, 1989; Condon, Weisberg, & Chambers, 2011). We also compared the predictive capacity of slope and topographic position index. Slope was the best fit, in agreement with Ziffer-Berger, Weisberg, Cablk, and Osem (2014) who found a weak correlation between A. tridentata and the topographic position index. Thus, only SLOPE was included in the models. All final climate, topographic, and fire (except FO) covariates were standardized to have a mean of zero and standard deviation of 1. We quantify the scale at which climatic, topographic, and anthropogenic variables operate in our study area in Supporting Information S6.

2.6 | Hurdle spatial models

Both occurrence and cover of a plant species determine its distribution in a landscape and can be governed by different environmental conditions (Holt, Gaston, & Fangliang, 2002). Thus, modeling occurrence and cover separately can produce substantially different results (Ospina & Ferrari, 2012) and contribute to misleading or divergent interpretations. To account for differences between occurrence and cover, we used hurdle spatial models to explore the effect of covariates on both response variables. Hurdle models permit modeling of distribution (presence and absence) and cover of plant species in an integrated framework (Irvine, Rodhouse, & Keren, 2016). The hurdle model (Cragg, 1971) is a two-component model able to accommodate two different spatial processes and is often used to fit data coming from two distributions (Potts & Elith, 2006; Tarbox, Fiestas, & Caughlin, 2018). Our model includes a Bernoulli process, representing sagebrush occurrence, that generates zeros and ones, where zero corresponds to absence and one corresponds to presence. If sagebrush does occur, we represent sagebrush cover with a Beta process that generates continuous numeric values between zero and one. We followed the notation used in Blangiardo and Cameletti (2015) to define the two hurdle model components as follows:

Let $y_o^i$ represent the observed sagebrush occurrence at location $i$:

$$y_o^i = \begin{cases} 1, & \text{if sagebrush does occur} \\ 0, & \text{otherwise} \end{cases}$$

and $y_c^i$ the observed sagebrush cover at location $i$:

$$y_c^i = \begin{cases} \text{NA}, & \text{if sagebrush does not occur at location } i \\ \text{Sagebrush cover at location } i, & \text{otherwise} \end{cases}$$

where $y_o^i \sim \text{Bernoulli}(x)$ and $y_c^i \sim \text{Beta}(a,b)$.

Plant occurrence and cover data are sensitive to residual spatial autocorrelation (i.e., model residuals at nearby locations are not independent; Tredennick et al., 2016). Integrated nested Laplace approximation (INLA) is a computationally efficient method for fitting complex models while accounting for spatial dependence of residuals and dealing with multiple likelihoods (Martins, Simpson, Lindgren, & Rue, 2013;
Rue, Martino, & Chopin, 2009). INLA was designed for fitting complex spatial models to nonnormally distributed response variables, including binary, count, and proportional data. INLA’s modeling framework builds off of link functions developed for generalized linear models, such as the logit-link, to map probability values to the entire real line. In the models proposed in our study, we apply the logit-link to model mean occupancy ($\mu^O_i$) and mean cover ($\mu^C_i$), at a site $i$, as linear functions of covariates and a spatial random effect term:

$$\logit(\mu^O_i) = a^O_i + \beta^O_{fire} + \beta^O_{treatment} + \sum_{j=1}^3 a^O_{ij} y_j + \phi^O_i$$ (1)

$$\logit(\mu^C_i) = a^C_i + \beta^C_{fire} + \beta^C_{treatment} + \sum_{j=1}^3 a^C_{ij} y_j + \xi_i$$ (2)

In Equations (1) and (2), $\mu^O_i$ represents the probability of a Bernoulli outcome, and $\mu^C_i$ represents the expected value of a Beta-distributed random variable. Additive parameters include intercepts of sagebrush occurrence and cover ($a^O_i$ and $a^C_i$), slope terms for the effect of fire ($\beta^O_j$ and $\beta^C_j$), treatment effects for the $k$th level of LTDL-derived restoration treatments ($\beta^O_{treatment}$ and $\beta^C_{treatment}$), and slopes for the $h$th climate and topographic covariate ($\omega^O_h$ and $\omega^C_h$). In addition, we included a spatial random effect ($\phi_i$) to account for residual autocorrelation in occupancy and cover. The shared spatial random effect for occupancy represents the correlation between the occupancy and cover components of the model.

The models were implemented in the R programming language using the R-INLA package (Martino & Rue, 2010). INLA’s approach to spatial modeling requires determining spatial thresholds for conditional independence between data points (Lindgren, Rue, & Lindström, 2011). We present the development of our model structure, including setting values for spatial thresholds, in Supporting Information S2. The code to reproduce the models is available on ScienceBase at https://doi.org/10.5066/P9NQNHH41 (Requena-Mullor, Maguire, Shinneman, & Caughlin, 2019).

We used five combinations of predictors to explore the effect of the fire attributes and restoration treatments on sagebrush occurrence and cover (Table 1). All the models included the two climate and topographic covariates, two intercepts (i.e., one per each model component), and a shared spatial random effect. Thus, our models accounted for covariation between sagebrush occurrence and cover, as such covariation patterns have been recognized for having important consequences in ecological modeling (Evans & Holsinger, 2012; Warton et al., 2015). We evaluated each of the attributes related to fire history in separate models (i.e., FO-model, occurrence of any fire during the study period; NOF-model, total NOF) and then selected the best performing fire covariate to include in a model with restoration treatment. We used the default priors provided by the R-INLA package for most parameters (Martino & Rue, 2010). The exception was the parameter representing correlation between spatial random effects for sagebrush cover and occurrence ($\phi$) in Equation (2). We assigned a weakly informative prior distribution to the correlation parameter, as a normal distribution with a mean of one and a precision of 10, to represent our expectation that cover and occurrence are likely to be positively correlated.

We evaluated the hurdle models by computing the mean absolute error (MAE) as a measure of fit (Liu, White, & Newell, 2011). The MAE measures the average magnitude of the absolute differences between the observed and fitted values, without considering their direction. This metric measures model prediction error in units of the variable of interest and, therefore, it is easy to interpret. To jointly assess the Bernoulli and Beta processes, we multiplied the fitted probability of occurrence by the sagebrush cover predictions and then evaluated these computed values of sagebrush cover relative to observed values. Additionally, we separately assessed predictive accuracy by calculating the Log-Loss (Phillips & Dudík, 2008) for

### Table 1: Hurdle spatial model evaluation using the test data.

The global performance of the hurdle models was evaluated using the mean absolute error (MAE), the Bernoulli component was assessed based on the Log-Loss, and the Beta component was evaluated based on MAE. Lower values of these model selection indices indicate better model performance. Performance measures were averaged throughout the seven data partitions (i.e., the folds yielded by the spatial blocking, see Supporting Information S3). A performance measure with letters in parentheses indicates that the performance was significantly better at a 0.05/4 level of confidence compared to the models indicated by the letters (Table S4; Figure S4). Models with no letters in parentheses mean that they were not significantly better than other competing models. We applied the Bonferroni correction to deal with the problem of multiple comparisons.

<table>
<thead>
<tr>
<th>Joint assessment</th>
<th>Bernoulli component</th>
<th>Beta component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>MAE</td>
<td>Log-Loss</td>
</tr>
<tr>
<td>aFO-model</td>
<td>0.0926 ±(0.008 SE)</td>
<td>0.6836 ±(0.08 SE)</td>
</tr>
<tr>
<td>bNOF-model</td>
<td>0.0926 ±(0.008 SE)</td>
<td>0.6837 ±(0.08 SE)</td>
</tr>
<tr>
<td>cNoFireVar-model</td>
<td>0.0934 ±(0.008 SE)</td>
<td>0.6835 ±(0.08 SE)</td>
</tr>
<tr>
<td>dTREAT-model</td>
<td>0.0832 ±(0.012 SE) [a,b,c]</td>
<td>0.5050 ±(0.04 SE) [a,b,c]</td>
</tr>
<tr>
<td>eNOF + TREAT-model</td>
<td>0.0825 ±(0.012 SE) [a,b,c]</td>
<td>0.5063 ±(0.04 SE) [a,b,c]</td>
</tr>
</tbody>
</table>

Note: NoFireVar-model included topography and climate covariates only. The lowest and highest values are shown in bold and underlined, respectively.

Abbreviations: FO, fire occurrence; NOF, number of fires; TREAT, restoration treatment.
sagebrush occurrence and the MAE for cover. Log-Loss measures the uncertainty of fitted probabilities by comparing them to observed presences. We used k-fold cross-validation of the sampled locations to test model fit to the data. A common issue when evaluating SDMs using cross-validation is that training and validation data randomly selected from nearby locations may be dependent because of spatial autocorrelation. Consequently, prediction errors can be underestimated, and may mislead model selection (Roberts et al., 2017). To account for spatial dependencies in our data, we adopted the spatial blocking strategy described by Roberts et al. (2017). This spatial blocking approach entails splitting the data into spatially independent blocks and then, sampling training and testing data from these folds. The spatial blocking approach forces more spatially distant records to serve as validation data, thus decreasing spatial dependence and reducing underestimation of prediction errors (Trachsel & Telford, 2016). See Supporting Information S3 for details. We used the test data to compute Log-Loss and MAE as performance metrics. To check if both performance measures (i.e., MAE and Log-Loss) were significantly different across the models, we explored the probability distributions of the normalized pairwise differences for each measure by a bootstrap method (10,000 replicates) and calculated which of these differences were significant using the Bonferroni correction (Table S4; Figure S4).

3 | RESULTS

Overall, models that included either fire, restoration treatment, or both of these anthropogenic covariates together provided better fits than models with just climate and topography. Considering the global performance of the hurdle models for both big sagebrush cover and occurrence (i.e., the ‘Joint assessments’ column in Table 1) the best fit included climate, topography, NOF, and restoration treatment. Without consideration of treatment effects, the joint assessment of the NOF-model showed the next highest goodness of fit, while the model that included climate and topography but no fire attributes yielded the lowest predictive capacity. Results were somewhat similar when evaluating the fit of the big sagebrush occurrence models separately from cover (Bernoulli component column in Table 1), as the best fit included both restoration treatment and NOF as predictor variables. However, without restoration treatment, there were no significant differences between big sagebrush occurrence models with and without fire attributes. Assessing model fit for big sagebrush cover separately from occurrence suggests potentially different relationships (Beta component column in Table 1), as the FO-model (fire occurrence) provided the best fit, and models that considered both fire and restoration treatment did not provide improvement. As a final check of the hurdle model, we also compared the performance of the NOF + TREAT-model by modeling both Bernoulli and Beta processes separately. Both of these separately fit models had higher values for Log-Loss and MAE (0.6836 ±0.08 SE) and 0.1149 ±0.007 SE, respectively) and, therefore, a lower prediction capacity than the joint model that simultaneously fit Beta and Bernoulli processes.

Scaled odds ratio = \( (e^{\beta} - 1) \times 100 \)  

where \( \beta \) and \( e^{\beta} \) represents a regression parameter estimated in the model and the change in the odds ratio when the predictor increases by 1 SD or the treatment factor changes from one level to another, respectively (Quinn & Keough, 2002). For example, an odds ratio of 0.553 (see the effect of FO on big sagebrush cover in Figure 3) represents a decrease of 44.7% in cover (i.e., \( (0.553 - 1) \times 100 \)), and an odds ratio of 1.234 (see the effect of Treatment on big sagebrush

![FIGURE 3](image-url)
occurrence) represents an increase of 23.4% in occurrence (i.e., \((1.234 - 1) \times 100\)).

There was large variability among fire attribute effect sizes (Figure 3). FO and NOF had negative effects on both big sagebrush occurrence and cover, but with considerably greater effects on cover. Models predicted that increasing from zero to at least one fire (i.e., FO) during the 36 year period of record would decrease big sagebrush occurrence by 0.5% (95% CI: −2.2%, 0.1%). The effect of 

fires on big sagebrush cover was much larger, with a predicted decrease in cover of 44.7% (95% CI: −47.9%, −41.3%) after increasing from zero to at least one fire during the study period. When one fire occurred (i.e., the mean of NOF + 1 SD) models predicted that big sagebrush occurrence would decrease by 1.2% (95% CI: −6.9%, 0.6%) and cover by 23.9% (95% CI: −26.6%, −21.8%). Restoration treatments increased the probability of big sagebrush occurrence, but exhibited higher uncertainty than any other parameter. The model predicted that the probability of occurrence increases by 23.4% (95% CI: 1.5%, 50%) when the site is treated and decreases by 19.0% (95% CI: −33.4%, −1.7%) if not. However, restoration treatment had no discernable effect on sagebrush cover.

A comparison between the influence of climate, topographic, and anthropogenic variables on big sagebrush is shown in Figure 4.

![Sagebrush occurrence and cover posterior distributions](image1)

**FIGURE 4** Posterior distributions of effect sizes for all covariates in best fit models. (a) Mean annual temperature; (b) cumulative annual precipitation; (c) slope; (d) fire occurrence; (e) number of fires; (f) treatment; (g) no-treatment. Effect sizes are shown in the response scale for big sagebrush occurrence and cover. Values lower than 1 indicate a negative effect on the response and values greater than 1 indicate a positive effect.

Both temperature and precipitation had negative effects, with a predicted decrease in occurrence of 59.1% and 60.7% when these variables increase by 3°C and 216 mm per year, respectively. SLOPE had a negative effect, and its effect size was greater than those of the fire variables. In contrast, climate and SLOPE had a weak influence on big sagebrush cover, with effect sizes smaller than that of the fire attributes (Figure 4). For example, our models predict that the occurrence of a single fire has a larger effect on big sagebrush cover than the effect of changing from minimum to maximum mean annual temperatures recorded across our sagebrush sites (Figure 5).

It is important to note here that we used ‘primary’ climate covariates to capture effects of climate in the Great Basin across multiple years. However, more nuanced covariates, such as seasonal or time-lag covariates, could yield different results (see for example Germino & Reinhardt, 2014; Tredennick et al., 2016).

Finally, the posterior estimates for the scale parameter \(k\) and the spatial variance \(\sigma^2\) were 1.50 (95% CI: 1.04, 1.94) and 0.265 (95% CI: 0.09, 0.47), respectively. The parameter \(k\) is related to the empirical range \(r\), that is, the distance at which the spatial correlation is 0.1 (Lindgren et al., 2011), which was 219.31 km for the sampled data. This distance suggests that the data are characterized by moderate spatial correlation (the maximum distance between two points was equal to 1,315.7 km, maximum mean distance for all the points is 927.4 km ± 173.3 SD). The posterior mean of the scaling parameter \(\phi\) for the shared spatial effect was 1.02 (95% CI: 0.43, 1.62), confirming the spatial correlation between sagebrush occurrence and cover. Refer to Supporting Information S2 for more detailed descriptions of these parameters.

**FIGURE 5** The predicted effect of the fire occurrence on sagebrush cover relative to the effect of climate, shown here using minimum (gray) and maximum (black) values of mean annual temperature. The logistic regression curves represent the relationship between fire occurrence (binary) and sagebrush cover. Dashed lines represent the 95% CI

4 | DISCUSSION

Species distribution modeling has emerged as a way to predict spatial patterns of species abundance over large areas; however, most SDMs are based on environmental variables that do not reflect the fine-scale impacts of anthropogenic disturbance. Our results demonstrate that including direct measurements of anthropogenic impacts in regional-scale SDMs can improve our understanding of species...
distributions. Our study species, big sagebrush, is a keystone species in the imperiled sagebrush steppe ecosystem (Miller et al., 2011). Prior studies have emphasized the role of climate and topography in determining sagebrush species distributions (Kleinhekselink & Adler, 2018; Schlaepfer, Lauenroth, & Bradford, 2012; Still & Richardson, 2015; Tredennick et al., 2016). However, sagebrush populations, especially in the Great Basin, are declining due to increases in fire frequency and are often additionally altered by intensive restoration efforts. Our spatially explicit SDMs helped to clarify and quantify how such anthropogenic factors have influenced big sagebrush distribution across the Great Basin. Models that included restoration treatment and fire history outperformed models with climate and topography alone. Indeed, the effect sizes of some of the fire-related variables were comparable to the effect sizes of climate variables, highlighting the importance of local-scale disturbance for predicting regional-scale abundance of big sagebrush. We also found that the inclusion of a restoration treatment variable significantly improved predictive accuracy of our models for big sagebrush occurrence, despite comprising decades of disparate restoration activities.

Climate-based SDMs and process-based models are increasingly used to forecast extinction risk and restoration outcomes (Shriver et al., 2018; Steen, Sofaer, Skagen, Ray, & Noon, 2017), and our work demonstrates the value of directly incorporating historical data on anthropogenic disturbance into these models to enhance their predictive capacity. For instance, by specifically examining the range of mean annual temperatures using minimum and maximum values among sagebrush occurrence locations, we also found that climate can differentially influence the predictions for big sagebrush abundance based on fire attributes (Figure 5). The prediction that sagebrush retains greater cover after fire in settings with colder mean annual temperatures supports recent classifications of sagebrush ecosystems across climate and soil gradients, in which sagebrush located in cooler and more mesic environments are generally considered more resilient after fire and more resistance to nonnative grass invasion (Chambers et al., 2014). Such results also reinforce findings that suggest successful restoration of big sagebrush after fire will benefit from climate-based adaptive management strategies (Shriver et al., 2018).

Altered fire regimes, including increased frequency and intensity of wildfire, are a consequence of climate change that is expected to impact numerous ecosystems around the globe (Robinson, 2009). However, fire history varies at fine scales across landscapes, and is perhaps a reason why SDMs that have incorporated fire attributes to predict species occurrence have reached different conclusions regarding the utility of fire attributes (Crimmins et al., 2014; Sherrill & Romme, 2012; Tucker et al., 2012). We expected big sagebrush occurrence and cover to exhibit a strong negative relationship with historical fire, as big sagebrush is considered vulnerable to increased fire frequencies due to its slow regeneration time and limited ability to compete with fire-adapted invasive grasses (Chambers, Roundy, Blank, Meyer, & Whittaker, 2007). Although we did find strong impacts of fire in general, its importance depended both on particular fire attributes being tested and whether big sagebrush occurrence or cover was the response variable (Figure 3). For big sagebrush occurrence, NOF was the best-performing fire attribute, while sagebrush cover was most strongly related to FO (i.e., whether a fire ever occurred at the site within the 35 year period examined). One potential explanation for these unique relationships is that, while a single fire is sufficient to reduce the cover of sagebrush relative to unburned stands over the duration of our 36 year fire history, multiple fires may be necessary to completely remove big sagebrush from a particular location. These results are corroborated by field studies of post-fire recovery of sagebrush in less-invaded landscapes. For instance, Shinneman and McIlroy (2016) found that a single fire was sufficient to keep big sagebrush communities from reaching preburn levels of cover for several decades on average, even though individual sagebrush plants often re-established within a few years. Regardless, by demonstrating that specific fire attributes vary in importance depending on the response variable being considered, these results also provide an explanation for why previous SDMs revealed no or varying effects of fire, given the different fire metrics, ecosystems, and species responses being tested among studies.

Along with the impacts of fire on species distribution, evaluating the influence of past restoration efforts is a critical research need (Brudvig et al., 2017). In the Great Basin, big sagebrush has been the focus of restoration efforts in recent decades, costing tens of millions of dollars annually (Arkle et al., 2014). A major challenge in integrating regional-scale restoration history into SDMs is that land management techniques change over time and implementation of recommended techniques varies between management units (Copeland et al., 2018). Despite this historical and spatial variability, we found that a relatively simple metric—whether or not a particular site had ever been part of a restoration treatment—had a significant positive effect on sagebrush occurrence. Indeed, for big sagebrush occurrence, the effect size of restoration treatment was comparable to the effect size of slope, albeit with much greater uncertainty. We suggest that integrating data on restoration history with other environmental variables is a worthwhile endeavor, even when land management records contain ambiguous information.

In contrast to sagebrush occurrence, we did not find a strong impact of restoration on sagebrush cover. We highlight three possible explanations for this apparent discrepancy. First, our treatment variable represents a range of restoration activities with different objectives, some of which are more likely to restore big sagebrush cover than others (Table S1). Second, restoration may increase the odds of sagebrush occurrence, but the high rates of seeding and planting failure and lengthy regeneration time required for big sagebrush recovery suggest that restoration efforts may require more time or innovative, climate-adapted restoration approaches to significantly increase cover on the landscape (Arkle et al., 2014; Beyers, 2004; Knutson et al., 2014; Shriver et al., 2018). Third, restoration treatments are not randomly placed, and land managers often select the most degraded landscapes for treatment where the likelihood of native species recovery is low (Pilliod, Welty, & Toews, 2017), while leaving relatively intact landscapes to recover without intervention, resulting in bias when
active restoration is compared to natural regeneration (Reid, Fagan, & Zahawi, 2018). Collectively, our results speak to the need for long-term demographic monitoring of restored plots to assess treatment effects, analyses that disaggregate restoration treatment into specific management actions, and the need for statistical analyses that more rigorously evaluate restoration impacts (Jones & Lewis, 2015; Pilliod et al., 2017).

While climate and anthropogenic covariates had different effects on big sagebrush occurrence and cover, there was a positive correlation in spatial pattern between the two response variables. One explanation for these differences is that anthropogenic factors complicate the relationship between environmental suitability and abundance (Dallas & Hastings, 2018; Estrada & Arroyo, 2012). In an undisturbed environment, we might expect strong positive correlations between probability of occurrence and cover, with the most suitable sites occupied at carrying capacity and the worst sites unoccupied. However, sagebrush steppe and shrublands are ecosystems at disequilibrium in the Great Basin, in which some populations are declining due to interactions with invasive species (Condon et al., 2011) and stochastic disturbance events, such as wildfires (Shinneman & McIlroy, 2016), while other populations are recovering from past disturbance, sometimes enhanced by restoration treatments. These local-scale human-induced factors may undermine the expected relationship between sagebrush distribution and climate. Climatically driven genetic differences of big sagebrush could further accentuate the disparity between the effects of climate and anthropogenic variables. Such genetic differences affect both sagebrush survival (Chaney, Richardson, & Germino, 2017) and its post-fire recovery (Brabec, Germino, & Richardson, 2016). In our study, different big sagebrush subspecies were modeled collectively (at the species level), but the subordinate taxa have been shown to respond differently to climate (Doherty et al., 2017) and fire (Brabec et al., 2016). Altogether, the degree to which habitat suitability predicted from occupancy data is correlated with abundance data is currently under debate in ecology (Dallas & Hastings, 2018; Weber, Stevens, Diniz-Filho, & Grelle, 2017). Although the causal mechanisms that might explain the disparity between modeled occupancy and cover are likely complex, our results emphasize the value of a joint-modeling approach that simultaneously accounts for both metrics of species abundance.

In our study, the Land Treatment Digital Library (Pilliod & Welty, 2013) enabled retrospective analyses of effects of land treatment and restoration on the sagebrush biome (Pilliod et al., 2017). As similar land management datasets emerge from other regions, we anticipate increasing potential to integrate anthropogenic factors into SDMs. For example, Mexico's National Plan for Ecosystem Restoration parallels many regional-scale restoration plans in Latin America (Tobon et al., 2017), and includes extensive documentation of the restoration activities undertaken (Méndez-Toribio, Martínez-Garza, Cecon, & Guariguata, 2018). While cadastral data (mapped property boundaries) may not include direct information on land management, these data provide opportunities to infer how landholder decisions may alter natural dynamics (Tarbox et al., 2018). Examples of open access cadastral data include the Rural Environmental Cadastre of Brazil (www.projects.worldbank.org) and the SIGPAC system of Spain, which integrate spatially explicit information of agricultural lands (www.fega.es). In addition to land management data, remote sensing platforms can provide spatially explicit data on historical disturbances, including wildfire (Davies, Ilavajhala, Wong, & Justice, 2009). In particular, Landsat satellite imagery has been successfully used to estimate a range of proxies for fire attributes at a relatively fine-scale, including fire severity (Eidenshink et al., 2007; Escuin, Navarro, & Fernández, 2008), active fire detection (Schroeder et al., 2016), and assessment of fire-affected areas (Chen, Moriya, Sakai, Kobayashi, & Cao, 2016). Altogether, these novel data sources provide a wealth of opportunities to quantify how anthropogenic factors influence communities and populations over regional scales.

While we applied data on sagebrush abundance and regional-scale covariates to model species distributions, an alternative approach could be to model the ecological processes (e.g., biogeochemical and hydrological cycles) that lead to changing distributions across space and time. For example, dynamic global vegetation models (DGVM) are increasingly being used to explore relationships between species distributions, fire, and climate variability, including in the sagebrush biome (e.g., Renwick et al., 2017). Complementary models for population dynamics also have the potential to forecast outcomes of restoration from patches to landscapes (Caughlin, Elliott, & Lichstein, 2016). The downside to these process-based models is that they require complex parameterization of plant ecophysiological traits and soil properties that influence establishment, growth, mortality, and competition among species, not all of which are well-quantified or available in the literature. In contrast, empirically derived SDMs typically require far fewer parameters and, therefore, fewer assumptions. While our models are static and correlative, we anticipate that they could inform dynamic vegetation models. For example, our logit-link models could enumerate critical thresholds needed as input for mathematically tractable models that explain alternative stable states in vegetation cover (Batllori, Ackery, & Moritz, 2015; D’Odorico, Laio, Ridolfi, Weissing, & DeAngelis, 2006).

A relatively simple addition to our SDM framework could be to jointly model the distribution of sagebrush and co-occurring species. A large body of evidence has demonstrated negative relationships between re-establishment of big sagebrush and invasive species, particularly B. tectorum, following fire (Chambers et al., 2007; Condon et al., 2011; West & York, 2002). Future SDM research in similar ecosystems could use our approach to model co-occurrence between key species and quantify how fire and restoration alter outcomes of interactions between species. Furthermore, as climate change will likely make sagebrush restoration increasingly difficult in the Great Basin (Shriver et al., 2018), our approach can be used to explore the interaction between climate and anthropogenic factors under future scenarios, thus informing land managers where restoration activities might be the most impactful.
Species distributions are increasingly affected by the growing scale and significance of human influence on natural processes. Previous studies have successfully incorporated anthropogenically derived predictors into multiscale analysis of species distributions (Austin, Thomas, Houston, & Thompson, 1996; Osborne, Alonso, & Bryant, 2001), including influence of roads, crops, urban settlements, and infrastructures (Mangiacotti et al., 2013; Suárez-Seoane, Osborne, & Alonso, 2002; Veran et al., 2016; Zuckerberg, Fink, Sorte, Hochachka, & Kelling, 2016) or coupled with political boundaries or population density indicators (Cuming, 2002; Greve et al., 2011; Lippitt et al., 2008). Our research included disturbance and anthropogenic factors different than those used previously in regional-scale SDMs for big sagebrush, and it improved our understanding of key dynamics affecting this keystone species in an imperiled biome. Although climate is recognized for exerting dominant control over species distributions at a regional scale, the influence of local-scale disturbance factors was nearly comparable in importance to that of climate and enhanced regional-scale predictive accuracy of both occurrence probability and abundance. For restoration practitioners seeking more realistic SDMs, our approach provides a method to unravel how both human-induced disturbance and restoration efforts scale up from a site-to-regional scale, and may help to better predict restoration success across scales. We further suggest that development of regional and global databases of land use and disturbance, coupled with additional methodological improvements to SDMs, will enhance the functionality of spatially predictive species modeling across multiple scales.

ACKNOWLEDGEMENTS

We would like to thank Josh Schraiber, Brianna McHorse, and Justin Welty for their help in developing the model framework and collecting fire data. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. This publication was made possible by the NSF Idaho EPSCoR Program under award number OIA-1757324 and by the U.S. Geological Survey. The models were run by using the High Performance Computing cluster of the Andalusian Computing Center, https://cica.es/servicios/supercomputacion/. We thank Andrii Zaiats, Cara Applestein, Anand Roopsind, and Cristina Barber for their helpful comments on the manuscript. Note: the term ‘Anthropocene’ is not formally recognized by the US Geological Survey as a description of geologic time and we use the term informally in this manuscript. The origins of the term are discussed by Zalasiewicz et al., (2008).

ORCID

Juan M. Requena-Mullor https://orcid.org/0000-0002-5120-7947
Kaitlin C. Maguire https://orcid.org/0000-0001-8193-2384
Douglas J. Shinneman https://orcid.org/0000-0002-4909-5181
Timothy Trevor Caughlin https://orcid.org/0000-0001-6752-2055

REFERENCES


biome. U.S. Geological Survey data release. https://doi.org/10.5066/P9NQNHH1


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.