esa

ECOSPHERE

Forecasting an invasive species' distribution with global distribution data, local data, and physiological information

CATHERINE S. JARNEVICH [1,†] NICHOLAS E. YOUNG [2] MARIAN TALBERT [3] AND COLIN TALBERT [1,3]

[1]*U.S. Geological Survey Fort Collins Science Center, 2150 Centre Ave Bldg C, Fort Collins, Colorado 80526 USA*
[2]*Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado 80523-1499 USA*
[3]*Department of Interior, North Central Climate Science Center, Colorado State University, Fort Collins, Colorado 80523 USA*

Citation: Jarnevich, C. S., N. E. Young, M. Talbert, and C. Talbert. 2018. Forecasting an invasive species' distribution with global distribution data, local data, and physiological information. Ecosphere 9(5):e02279. 10.1002/ecs2.2279

Abstract. Understanding invasive species distributions and potential invasions often requires broad-scale information on the environmental tolerances of the species. Further, resource managers are often faced with knowing these broad-scale relationships as well as nuanced environmental factors related to their landscape that influence where an invasive species occurs and potentially could occur. Using invasive buffelgrass (*Cenchrus ciliaris*), we developed global models and local models for Saguaro National Park, Arizona, USA, based on location records and literature on physiological tolerances to environmental factors to investigate whether environmental relationships of a species at a global scale are also important at local scales. In addition to correlative models with five commonly used algorithms, we also developed a model using a priori user-defined relationships between occurrence and environmental characteristics based on a literature review. All correlative models at both scales performed well based on statistical evaluations. The user-defined curves closely matched those produced by the correlative models, indicating that the correlative models may be capturing mechanisms driving the distribution of buffelgrass. Given climate projections for the region, both global and local models indicate that conditions at Saguaro National Park may become more suitable for buffelgrass. Combining global and local data with correlative models and physiological information provided a holistic approach to forecasting invasive species distributions.

Key words: buffelgrass; *Cenchrus ciliaris;* climate change; correlative model; habitat suitability; response curves; Saguaro National Park; scale; species distribution model; user-defined curve.

Received 19 January 2018; revised 17 April 2018; accepted 20 April 2018. Corresponding Editor: Debra P. C. Peters.
Copyright: © 2018 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
† **E-mail:** jarnevichc@usgs.gov

INTRODUCTION

To understand the effect of climate on a species' distribution, it is necessary to examine the species' response to climate across broad gradients, often at a continental or global scale (Parmesan and Yohe 2003, Seo et al. 2009, Franklin et al. 2013). Understanding broad-scale constraints on species distributions is important in the context of invasion ecology (Thuiller et al. 2005) and climate change (Pearson and Dawson 2003). Climate is often thought to control broad-scale distributions (Pearson et al. 2004) with impacts occurring across multiple scales. Examining global distributions can provide observation-based climate tolerance information for invasive species. Additionally, models that predict suitable habitat for invasive species perform better when including data from both native and invaded ranges (Jiménez-Valverde et al. 2011, Evangelista et al. 2016), as data across all geographic areas of occurrence capture the full extent of environmental conditions experienced by a species.

Conversely, resource managers perform their work at finer landscape scales, often focused

within ownership boundaries, requiring local-scale information to inform management activities. Factors controlling distribution at landscape scales may be different than those defining broad geographic patterns, with edaphic, topographically modulated climate, disturbance, and biotic factors determining species' distributions within these broad regions of climatic suitability (Pearson et al. 2004). Changing climate may impact local suitability, and predicting this impact requires broad information about a species' abiotic tolerances. This information could be derived from broad-scale models or literature on the physiological relationship of a species' with environmental conditions.

Buffelgrass (*Cenchrus ciliaris* L. syn *Pennisetum ciliare* (L.) Link) has been widely introduced globally as a key forage species for livestock (Marshall et al. 2012). Along with its wide introduction globally, it has become invasive in natural areas in many parts of the world, creating a tension between commercial value and maintaining natural ecosystems (Grechi et al. 2014). Given its economic importance, there have been many studies examining buffelgrass's climatic tolerances related to where it is able to persist and produce high biomass, which is likely linked to impact in introduced regions where it has become invasive (Appendix S1: Table S1). This information can be used to develop a priori relationships between environmental characteristics and buffelgrass occurrence. Species distribution models can also be used to identify relationships between species and environmental conditions. The appropriateness of correlative species distribution models to predict potential changes in distributions has been criticized (Fitzpatrick et al. 2007). They can be useful when informing management activities related to invasive species (e.g., Crall et al. 2013). Further, there are ways that these models can be validated as to the mechanisms underlying the predictions such as comparing ecophysiologically derived response curves to those derived via statistical correlations.

In this study, we investigate whether climatic relationships identified from a plant species' global distribution match those identified at a local landscape using climate data downscaled to reflect microclimate conditions to predict the local, fine-scale distribution. In addition, we develop a global distribution model by defining response curves based on ecophysiological tolerances reported in the literature to evaluate correlative model results. We also investigate whether species abundance at a local level differs in environmental space. We use these different analyses to hypothesize the distribution of buffelgrass within a discrete management unit, including current conditions and potential future regional climatic conditions.

Methods

Location data

Global extent.—We obtained 1237 records for buffelgrass from the Global Biodiversity Information Facility (GBIF) on 27 October 2015 (https://doi.org/10.15468/dl.x76i7f). We selected data that did not have known spatial issues, had coordinates within the identified country of observation, and with a date from 1 January 1950 or later. We chose to use the target background method suggested by Phillips et al. (2009) for use with museum data. Therefore, we downloaded records from GBIF on 27 October 2015 for the family Poaceae from the subset of countries and institutions for which we had buffelgrass locations (https://doi.org/10.15468/dl.njt6ad). We applied the same criteria to select records for the background, with the addition of the country and institution filter where we limited these to ones with buffelgrass records.

Local extent.—The 36,960 ha Saguaro National Park, located in southeastern Arizona, USA, is divided into two units with the city of Tucson in between (Fig. 1). The two units are the Rincon Mountain District (RMD) with elevations between 813 and 2641 m and the Tucson Mountain District (TMD) with elevations between 666 and 1420 m. The RMD is east of Tucson and extends to higher elevations than the TMD unit located west of Tucson. The Park falls within the Arizona Upland division of the Sonoran Desert which is the highest elevation division and has the most frequent hard frosts (Dimmitt 2000). The Park contains multiple vegetation types including desert scrub, desert grassland, oak woodland, and ponderosa pine forest.

Buffelgrass was first observed in the Park in 1989, and an active management program began in 1993 with manual removal expanding to herbicide application in 2005. Since then,

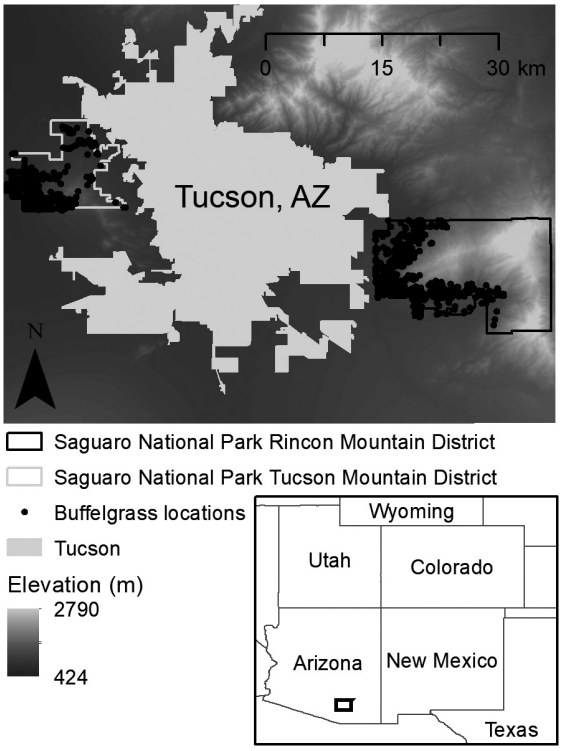


Fig. 1. Location of the two units comprising Saguaro National Park, highlighting buffelgrass occurrence locations.

approximately 200 ha have been treated in accessible locations. The Park has buffelgrass ground mapping data from 1997 to the present that includes location and estimated percent cover. There are also data from aerial surveys in 2007 and 2012, and these surveys cover areas of the Park that are difficult to access. We limited our analyses to the Park boundaries because adjacent areas have multiple land ownership with more varied surveys, include unknown data collection gaps, and include different environmental conditions (e.g., developed areas of Tucson, Arizona, USA, compared to natural habitat within the Park). We extracted information from georeferenced locations from 2005 forward due to spatial accuracy issues with earlier data. We were interested in the maximum possible percent cover for any pixel, so we extracted the maximum percent cover class value ever recorded to provide some control for active management impacts on cover. We analyzed these data with our local predictor variables to see whether there were any environmental differences between cover classes. The buffelgrass mapping data from the Park is very intensive, so we spatially filtered the data for species distribution model development in R, removing any locations within 100 m of each other using the rgdal (Bivand et al. 2016) and sp (Bivand et al. 2013) packages with R version 3.3.1 (R Core Team 2016). This filtering reduced the number of samples from 30,166 to 1600 locations. Because the Park conducted an aerial survey of the entire area in 2014 and thus believes they have captured all buffelgrass patches in their mapping, we treated the rest of the Park as buffelgrass absence and generated an equal number of random absence locations in raster cells without a presence location within the Park boundaries (i.e., $n = 1600$).

Predictor variables

Global extent.—Potential global predictor variables included climate. We obtained 19 bioclimatic variables at 30 arc second resolution from the WorldClim data set (Hijmans et al. 2005). We first selected predictors with biological relevance based on a literature review of buffelgrass growth limitations (Appendix S1: Table S1), followed by correlation filtering where we removed one of any pair with a Pearson, Spearman, or Kendall correlation coefficient >|0.7| to minimize collinearity issues (Dormann et al. 2013) using the Software for Assisted Habitat Modeling (SAHM) CovariateCorrelationAndSelection module (Morisette et al. 2013).

Local extent.—Local-scale predictors included climatic and topographic variables. We generated climate data with the ClimateNA v5.10 software package (Wang et al. 2016) using NED 10 (National Elevation Dataset at 1/3 arc second resolution or approximately 10 m) to downscale the climate data. ClimateNA uses statistical algorithms to relate elevation and climate to produce a statistically downscaled version of PRISM climate data (PRISM Climate Group 2012). The baseline data were PRISM climate normals for 1981 to 2010. We considered 13 climate predictors directly calculated from annual data and seasonal data. Interpolated precipitation data in the region do not capture the heterogeneity in most precipitation events, so we did not include these data at the local scale. We instead derived topographic variables from the NED 10 m elevation layer

using the R raster package terrain function with eight neighbors (Hijmans 2015), including eastness, northness, slope, topographic position index, topographic ruggedness index, and roughness. We used the same process to eliminate collinear variables for the local extent as we did for the global extent.

We were also interested in the potential influence of disturbance on the distribution of buffelgrass within the Park. As such, we created a second model, with presence and absence data limited to environmentally suitable locations as determined by the first model, using anthropogenic features as predictors including distance to trails, road density, and distance to roads.

Environmental space and buffelgrass abundance

We divided the location data in the Park into different bins related to recorded percent cover levels. These bins reflected different management priority classifications and included <1%, 1–10%, 10–50%, and >50% buffelgrass cover (Saguaro National Park staff, *personal communication*). Locations with <1% cover represent maintenance level for Park management. At 10% cover, these areas have enough biomass to carry a fire (P. Grissom, *personal communication*), and >50% is the cover threshold for aerial spraying that is identified in the Saguaro National Park management plan. We again selected the maximum cover ever reported when classifying a location to a bin. We generated scatter plots for the point locations using each pair of the different predictors used in the habitat suitability modeling using GGally (Schloerke et al. 2017) within R.

Model development

All models were created using the SAHM modules (Morisette et al. 2013) within the VisTrails software (Freire et al. 2006). We used all six model algorithms available within SAHM for the global model which include boosted regression trees (BRT; Elith et al. 2008), generalized linear models (GLM; Hosmer and Lemeshow 2000), Maxent (Phillips et al. 2006), multivariate adaptive regression splines (MARS; Elith and Leathwick 2007), and random forest (RF; Breiman 2001). For the global user-defined curve model (UDC), we drew response curves a priori based on information from the literature about buffelgrass environmental tolerances (Appendix S1: Table S1). Unlike the other five modeling modules that are statistically based, this module produces predicted habitat suitability based on a mechanistic understanding of the species. All response curves are multiplied together, which means that the predictor with the lowest relative probability will control the modeled probabilities. Thus, we considered both the shape of the curves and our understanding of relative importance of predictors in developing the UDC. While we felt confident in creating UDCs based on broad-scale climatic relationships, we only included this method at the global scale due to the limited landscape-scale information. While aerial surveys will not capture an individual plant in a pixel, the Park's assessment that they have captured most buffelgrass in combination with our analysis showing no relationship between cover and environmental conditions led us to treat these data as a comprehensive survey of buffelgrass occurrence. We therefore treated our randomly generated locations outside of patches as absence data and thus did not include Maxent at the local scale.

We used 10-fold cross-validation to assess model performance. We initially ran all models using the default settings in SAHM. For any models exhibiting poor performance (AUC < 0.7) or >0.05 difference between test and train AUC values, we optimized individual model parameter settings. For the global models, we tested RF node size from 50 to 500 by increments of 50 and selected a node size of 100 for model development. Various combinations of BRT learning rate and tree complexity were compared, and we identified a learning rate of 0.005 and tree complexity of 4 for the final model. Finally, we evaluated Maxent regularization values between one and ten and selected a regularization value of two for model development.

We assessed model performance using several common discrimination metrics including AUC, true skill statistic (TSS), correlation coefficient, and percent correctly classified. We also examined model calibration using the presence-only calibration plot (Phillips and Elith 2010) implemented in SAHM and calibration statistics testing bias, spread, and unexplained error (Pearce and Ferrier 2000). We examined variable importance, calculated by assessing the difference in AUC values when values for the predictor of interest were permutated between presence and

background locations while holding the other predictors constant. We then calculated relative importance of predictors by dividing these values by the sum of importance for all predictors for that model. We also produced environmental variable response curves for all models.

From the set of final optimized models, we selected the ones with TSS >0.4 that were well calibrated and ecologically plausible (e.g., response curves matched a priori expectations). We generated an ensemble of the selected models by taking an average of the probabilities and created a binary presence/absence map using a threshold of 0.5 to discretize predictions.

Climate impact assessment

We obtained climate normals derived from PRISM data for 1981–2010 and future climate predictions from 15 scenarios available in ClimateNA using the GMTED2010 30 arc second elevation data set (Danielson and Gesch 2011). We extracted predictions for the representative concentration pathway 8.5 for the year 2055 for each global circulation model. We calculated the same set of climatic factors used in the global models to place the Park climate conditions within the context of the response curves from the global models. We examined how Saguaro National Park's location on the global response curves might change with changing climate by examining the range of values found in each unit currently and in the future scenarios.

Results

Model files and results are available as a U.S. Geological Survey data release (Jarnevich 2018).

Distribution models

Global extent.—After correlation filtering, we retained four predictor variables in the global models that were related to observed physiological tolerances that included winter temperature and precipitation along with summer temperature and precipitation. The global models, except for the UDC, all performed reasonably well with cross-validation AUC values > 0.7 and TSS values > 0.38 (Table 1a).

The user-defined response curves that were based on our literature review matched the general pattern of those derived by the statistical models (Fig. 2a). In all cases, winter temperature was the most important predictor (Table 2a), with increasing suitability beginning between 5° and 10°C and generally leveling off by 20°C with some noise in the response curves. Summer precipitation and winter precipitation ranked second and third, respectively, in four of the models while switching order in the other two. Suitability sharply increased with a small amount of summer precipitation and then decreased starting with 250–400 mm to low suitability around 750 mm. Winter precipitation had high suitability with low precipitation, with decreasing suitability to around 600 mm where it leveled off. Mean

Table 1. Assessment metrics for each model algorithm.

Algorithm	AUC	TSS	CC	PCC	Calibration
(a)					
BRT	0.79 (0.81)	0.43 (0.45)	0.13 (0.13)	73 (72)	$P = 0$ ($P = 0$)
GLM	0.77 (0.77)	0.39 (0.39)	0.11 (0.12)	70 (70)	$P = 0$ ($P = 0$)
MARS	0.78 (0.79)	0.39 (0.42)	0.12 (0.13)	70 (70)	$P = 0$ ($P = 0$)
Maxent	0.80 (0.80)	0.42 (0.43)	0.13 (0.13)	72 (72)	$P = 0$ ($P = 0$)
RF	0.80 (0.80)	0.43 (0.43)	0.13 (0.13)	72 (71)	$P = 0$ ($P = 0$)
UDC	0.68	0.26	0.07	63	$P = 0$ ($P = 0$)
(b)					
BRT	0.87 (0.89)	0.57 (0.60)	0.65 (0.68)	78 (80)	$P = 0.48$ ($P = 4.4\text{e-}5$)
GLM	0.86 (0.86)	0.58 (0.58)	0.64 (0.64)	79 (79)	$P = 0.38$ ($P = 1$)
MARS	0.87 (0.87)	0.59 (0.60)	0.66 (0.67)	79 (80)	$P = 0.4$ ($P = 1$)
RF	0.80 (0.83)	0.42 (0.55)	0.57 (0.59)	71 (77)	$P = 0$ ($P = 0$)

Notes: Model algorithms are boosted regression trees (BRT), generalized linear model (GLM), multivariate adaptive regression splines (MARS), Maxent, random forest (RF), and user-defined curves (UDC) and were calculated for the cross-validation data (and training data in parentheses) for the (a) global models and (b) Saguaro National Park models including area under the receiver operating characteristic curve (AUC), true skill statistic (TSS), correlation coefficient (CC), percent correctly classified (PCC), and calibration *P* value for intercept is 0 and slope is 1.

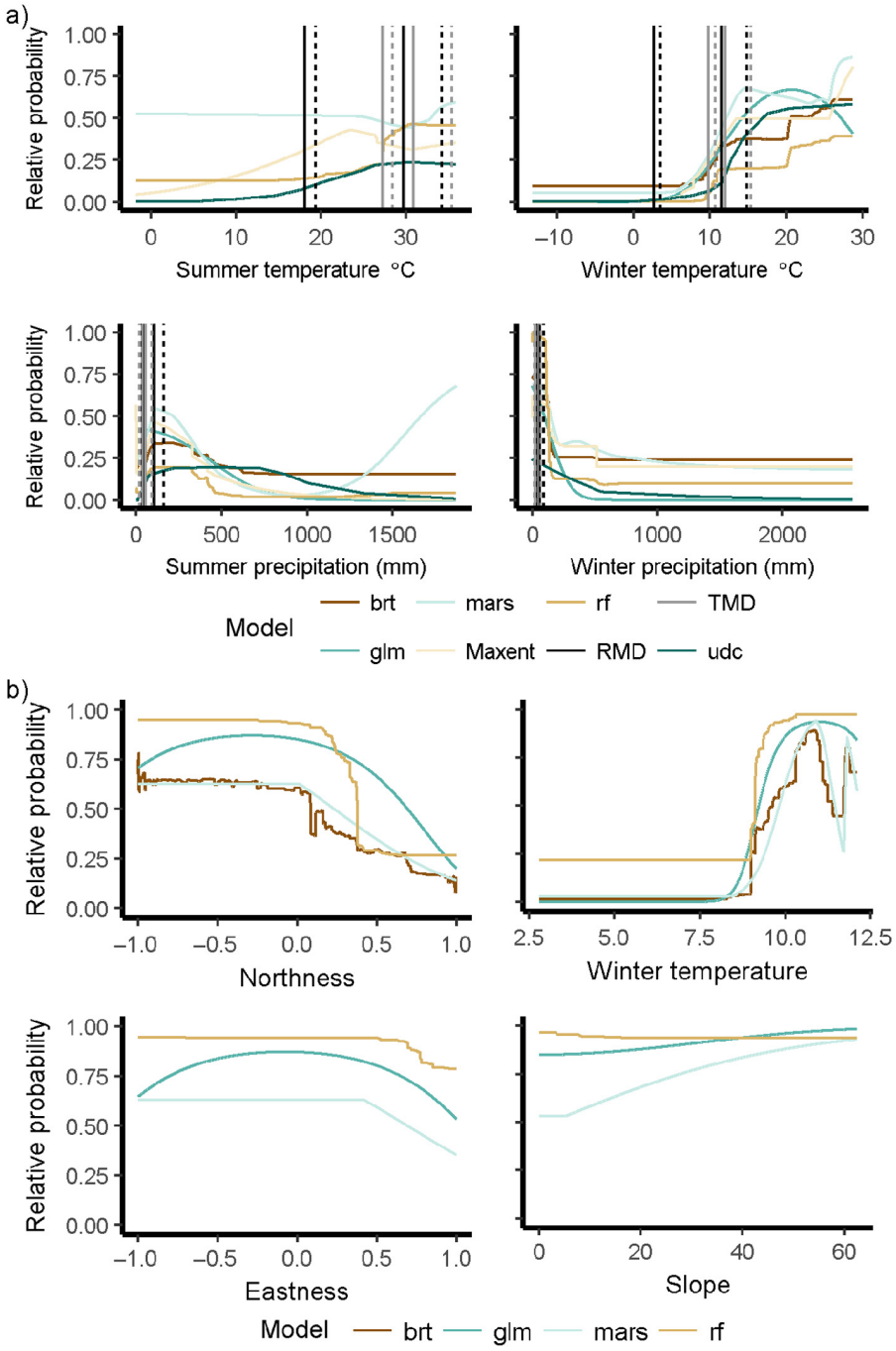


Fig. 2. Response curves for each predictor variable for the (a) global models and (b) local models. Each predictor includes a curve from any algorithm (boosted regression trees (BRT), generalized linear models (GLM), multivariate adaptive regression splines (MARS), and random forest [RF]) that retained the predictor during internal variable selection. Global models also include Maxent and user-defined curves (UDC) models. Global models include solid lines depicting the range of climate conditions for 1981–2010 and dotted lines depicting the range predicted by 15 different global circulation models for representative concentration pathway 8.5 for 2055, with Rincon Mountain District (RMD) in black and Tucson Mountain District (TMD) in gray.

Table 2. Relative variable importance expressed as a percent of the total contribution for each model algorithm† for the (a) global models and (b) Saguaro National Park models.

Predictor variable	BRT	GLM	MARS	Maxent	RF	UDC
(a)						
Summer temperature	n/a	n/a	3	6	9	0
Winter temperature	50	49	52	48	38	65
Summer precipitation	26	19	30	31	28	23
Winter precipitation	24	32	15	15	25	12
(b)						
Winter temperature	77	66	77	–	67	–
Eastness	n/a	8	2	–	4	–
Northness	23	24	17	–	30	–
Slope	n/a	2	4	–	0	–

Note: The en dash indicates when model algorithm was not used for the area.
† Boosted regression trees (BRT), generalized linear model (GLM), multivariate adaptive regression splines (MARS), Maxent, random forest (RF), and user-defined curves (UDC).

summer temperature was dropped from two of the three models with internal variable selection and contributed <10% to any of the other models.

The global spatial predictions showed general agreement across models (Fig. 3; Appendix S1: Fig. S3). Africa and Australia were highly suitable for buffelgrass along with southern Asia, southern North America, and portions of east and central South America. However, the UDC model only predicted the northern portion of Australia as suitable while all other models showed suitability extending further south.

Local extent.—We retained four predictor variables for the local-scale model that were related to buffelgrass tolerances, including winter temperature, eastness, northness, and slope. Different cover classes of buffelgrass did not occupy different components of environmental space (Appendix S1: Fig. S1). There were more high cover records at locations with lower average winter temperature, but this was likely due to sampling bias from the aerial surveys that covered higher elevation and difficult to access locations. Therefore, all models were created by pooling all occurrence data, regardless of cover level.

All local-scale models performed well according to assessment metrics, but RF and BRT were not well calibrated as indicated by *P*-values testing that the slope was 1 and intercept was 0 (Table 1b). Additionally, the response curves for BRT appeared overfit (Fig. 2b), so we dropped both of these models from the ensemble. Winter temperature had the greatest influence on model predictions, followed by northness (Table 2b). Suitability was greatest in areas with warmer winter temperature, beginning to decline around 10°C, and on south facing slopes with suitability decreasing as a slope became more northward facing. Eastness and slope had a combined contribution of <10%.

All local model spatial predictions show the majority of TMD being suitable for buffelgrass with only the western, lower elevation portion of RMD being suitable habitat (Appendix S1: Fig. S3). The BRT and MARS models predicted the northern portion of the Tucson Mountains around Wasson peak in TMD as less suitable for buffelgrass compared to the GLM and RF model.

The models developed using anthropogenic features had poor assessment metrics (e.g., AUC < 0.7). Given the poor performance of the models, we do not include any results output for these models.

Climate impact assessment

Mean winter temperature in Saguaro National Park between 1981 and 2010 ranged from 2.7° to 11.5°C in RMD and from 9.8° to 11.6°C in TMD. The higher elevations within the Park match conditions identified as unsuitable according to winter temperature values and global distribution models. All future scenarios resulted in an increase in the minimum mean winter temperature within the Park, resulting in a shift toward greater suitability for buffelgrass. Mean summer temperature in Saguaro National Park between 1981 and 2010 ranged from 18.1° to 29.7°C in RMD to 27.2° to 30.9°C in TMD. Again, all future scenarios resulted in a shift in the range toward warmer temperatures, following an increase in

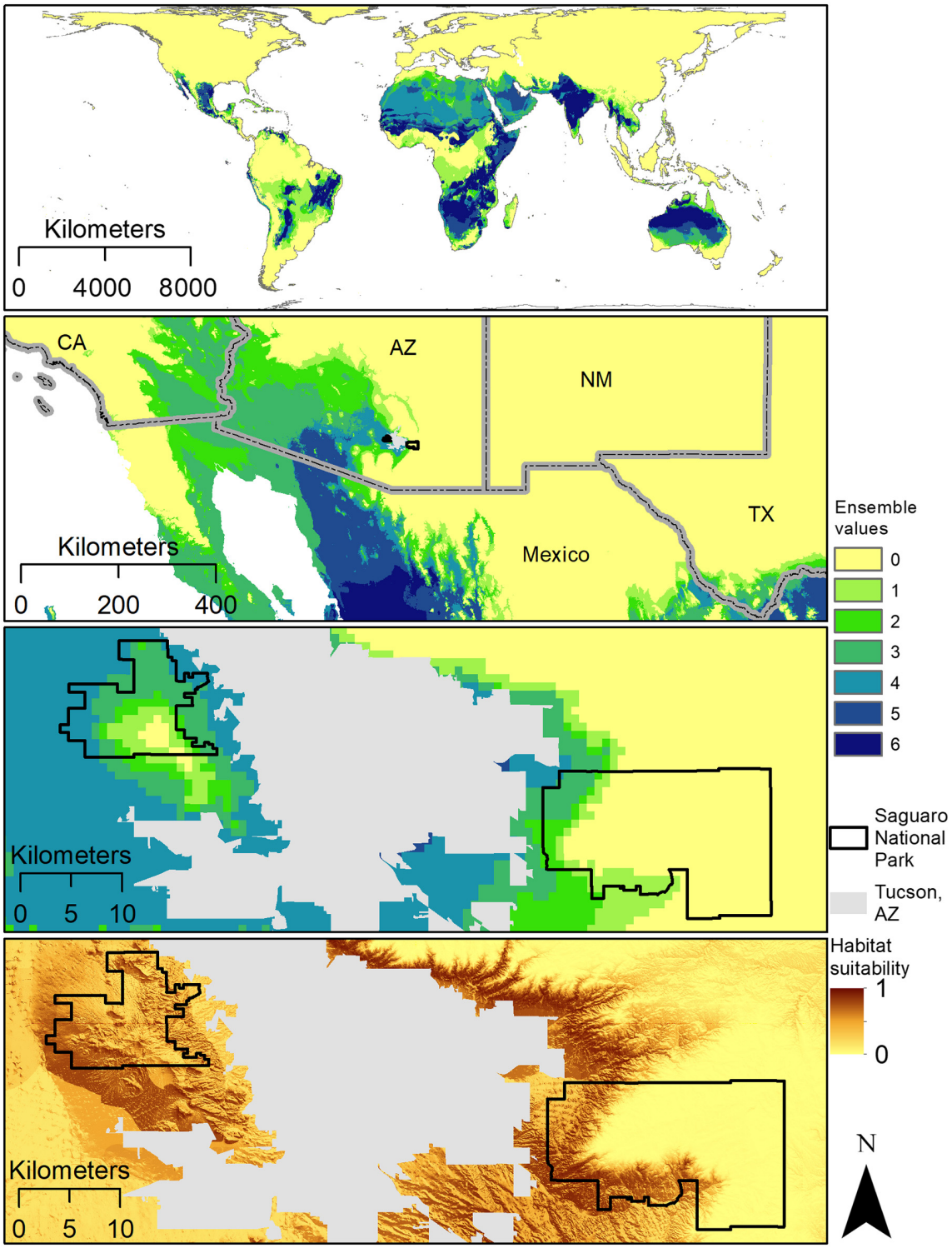


Fig. 3. Maps of the habitat suitability models including an ensemble of the global model, a regional view of the global ensemble, the global ensemble for the Tucson, Arizona, USA, area, and the local habitat suitability model for the Tucson, Arizona, USA, area.

predicted suitability on the global response curves (Fig. 2a; Appendix S1: Fig. S2).

For the Park, there was little difference in summer and winter 1981–2010 precipitation normals and in any of the future climate scenarios except for three scenarios showing an increase in summer precipitation and one scenario showing an increase in winter precipitation. These differences, however, were not noticeable when plotted against the range of precipitation found at the global scale (maximum within Park predicted as 158.7 mm by CanESM2 while global curves ranged to >2000 mm). Both current and future precipitation in the Park matched higher suitability on the global response curves.

Discussion

Species distribution models approached from both a correlative and process-based dichotomy produced similar results, lending confidence to predictions. In this case, our user-defined response curves based on a literature review of environmental conditions limiting buffelgrass growth showed a similar pattern to those derived from statistical models based on occurrence data related to the same set of climatic predictors.

Winter temperatures have been observed to control where buffelgrass can occur, predominate, and spread (Cox et al. 1988, Hussey and Bashaw 1996), and it has been shown to have a freezing tolerance around ~°C (Stair et al. 1998). Additionally, there is evidence that extremely high temperatures are lethal to buffelgrass (Cox et al. 1988, Barrera and Castellanos 2007); locations with these high temperatures did not occur in the study area. Optimal temperature for buffelgrass growth is around 30°C (Barrera and Castellanos 2007); therefore, our response curve for summer temperatures steadily increased to this temperature. Cox et al. (1988) observed that buffelgrass either dies or persists without domination or spread with >600 mm of precipitation during the growing season, and Rao et al. (1996) observed the greatest yield in India with 180–250 mm of rain. While in Mexico, the greatest observed biomass over a three-year period occurred in the wettest year (July to September monsoon precipitation of 358 mm; Martin-R et al. 1995). Buffelgrass has also been observed to predominate and spread where winter precipitation is <400 mm (Cox et al. 1988).

Marshall et al. (2012) compiled global distribution information for buffelgrass to country or state level from published or Internet sources. Looking at relative suitability from our models, most areas identified by Marshall et al. (2012) had at least some predicted suitability with the exception of Maritime Southeast Asia and Europe (although coastal Spain had higher relative suitability compared to other areas). Qualitative comparison with previous models for other regions such as Australia matched with our ensemble model and Martin et al. (2015) predicting lower suitability for northern coastal regions and southern Australia. Predictions for Sonora, Mexico, visually agreed with other predictions with relatively higher suitability in the west compared to the northeast (Arriaga et al. 2004).

Winter temperature was the largest contributing environmental variable for both the global and local models, suggesting that broad-scale environmental drivers for a species distribution persist at the local scale. Further, our local-scale models showed a similar response to winter temperature as the global models, which included a broader environmental gradient encompassed by global occurrences of buffelgrass. Winter temperature's relative importance increased at the local scale compared to the globe, but explained almost half the variation in all models. Other predictors in the two models differed as we included topography as a proxy for other climate data at the local scale rather than the same climate predictors we used at the global level due to the heterogeneity in precipitation in the region relative to the number of stations available to interpolate precipitation patterns. This follows the idea that climatic drivers at the global scale continue to be important at the local scale, although topography can mediate climate at the local scale (Austin and Van Niel 2011, Dobrowski 2011).

The data used to develop both global and local models were not collected to create species distribution models. Global Biodiversity Information Facility contains many spatial biases (Yesson et al. 2007), and we attempted to account for this using the target background approach. However, model relationships may be influenced by these biases. At the local scale, there may be detection issues, especially in the aerial survey zones of the Park. For invasive species that are still spreading, there are also issues related to absence data in

that a location may be an absence because the species has not yet arrived at that location. These models can still be useful in a search context and should be updated through an iterative process as new locations are detected (Crall et al. 2013).

Our results suggest that there is no evidence that measured environmental factors are limiting the abundance of buffelgrass across the Park, and thus, buffelgrass may become abundant anywhere with suitable habitat. These results are similar to those from the southeastern USA, finding no limitations of infilling of invaded locations by another invasive plant, *Microstegium vimineum* (Warren et al. 2013). Fine-scale studies along ecotones (e.g., across the winter temperature gradient from suitable to unsuitable conditions) might capture environmental tolerances related to abundance, but this was beyond the scope of this study. Additionally, there is evidence that soil properties may be related to abundance from a regional study in Mexico (Ibarra-Flores et al. 1995), but soil data currently available for the Park do not adequately capture the existing heterogeneity to be useful (e.g., finest scale SSURGO soil data available at 1:12,000 to 1:36,000 scale). Also, a previous model including soils did not find support for the importance of soil data (Ward 2003), hypothesized due to the coarse soil mapping unit. Additionally, the lack of performance of the anthropogenic predictor model could indicate that at this stage of the buffelgrass invasion within the Park, anthropogenic pathways are no longer driving the invasion. Anthropogenic features such as roads and trails can act as corridors of invasion (Hulme 2009), but at later stages of the invasion process, they may not be as important if the species is able to spread and invade in undisturbed locations as buffelgrass seems to do in this region (Olsson et al. 2012). Further, wind is an important dispersal agent (Stevens and Falk 2009), so once buffelgrass is established in an area, anthropogenic pathways may not be as limiting for further dispersal as in species with other kinds of dispersal mechanisms.

Conclusions

We developed global and local models of buffelgrass in Saguaro National Park, AZ, and evaluated how potential future climate may impact the distribution of the species in the Park. All global and local correlative models included winter temperature as the greatest contributor, emphasizing the importance of climate as a driver across scales. At the local scale, additional topological predictors were proxies for other climate predictors. Additionally, the model based on buffelgrass ecophysiological information had similar results to the statistical models, providing confidence in model predictions. Given our model results and the suite of climate projections available for the region, Saguaro National Park may become more suitable for buffelgrass in the future. Combining global and local occurrence and environmental data to construct correlative models while also developing an ecophysiological-based model provided a comprehensive approach that can be applied to other invasions to forecast invasive species distributions for a landscape.

Acknowledgments

Funding to support this work was provided by the Natural Resources Preservation Project (NRPP) and the U.S. Geological Survey Invasive Species Program. We thank Catherine Cullinane Thomas and two anonymous reviewers for comments on earlier versions of this manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

Arriaga, L., A. E. Castellanos, E. Moreno, and J. Alarcon. 2004. Potential ecological distribution of alien invasive species and risk assessment: a case study of buffel grass in arid regions of Mexico. Conservation Biology 18:1504–1514.

Austin, M. P., and K. P. Van Niel. 2011. Improving species distribution models for climate change studies: variable selection and scale. Journal of Biogeography 38:1–8.

Barrera, E. D. L., and A. E. Castellanos. 2007. High temperature effects on gas exchange for the invasive buffel grass (*Pennisetum ciliare* [L.] Link). Weed Biology and Management 7:128–131.

Bivand, R., T. Keitt, and B. Rowlingson. 2016. rgdal: bindings for the Geospatial Data Abstraction Library. R package version 1.2-5. https://CRAN.R-project.org/package=rgdal

Bivand, R. S., E. Pebesma, and V. Gomez-Rubio. 2013. Applied spatial data analysis with R. Second edition. Springer, New York, New York, USA.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

Cox, J. R., M. H. Martin-R, F. A. Ibarra-F, J. H. Fourie, J. F. G. Rethman, and D. G. Wilcox. 1988. The influence of climate and soils on the distribution of four African grasses. Journal of Range Management 41:127–139.

Crall, A. W., C. S. Jarnevich, B. Panke, N. Young, M. Renz, and J. Morisette. 2013. Using habitat suitability models to target invasive plant species surveys. Ecological Applications 23:60–72.

Danielson, J. J., and D. B. Gesch. 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey Open-File Report 2011–1073.

Dimmitt, M. A. 2000. Biomes and communities of the Sonoran Desert region. Page 18 *in* S. J. Phillips, editor. A natural history of the Sonoran Desert. Arizona-Sonora Desert Museum Press, Tucson, Arizona, USA.

Dobrowski, S. Z. 2011. A climatic basis for microrefugia: the influence of terrain on climate. Global Change Biology 17:1022–1035.

Dormann, C. F., et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:027–046.

Elith, J., and J. Leathwick. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions 13:265–275.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77:802–813.

Evangelista, P. H., N. E. Young, P. J. Schofield, and C. S. Jarnevich. 2016. Modeling suitable habitat of invasive red lionfish *Pterois volitans* (Linnaeus, 1758) in North and South America's coastal waters. Aquatic Invasions 11:313–326.

Fitzpatrick, M. C., J. F. Weltzin, N. J. Sanders, and R. R. Dunn. 2007. The biogeography of prediction error: Why does the introduced range of the fire ant over-predict its native range? Global Ecology and Biogeography 16:24–33.

Franklin, J., F. W. Davis, M. Ikegami, A. D. Syphard, L. E. Flint, A. L. Flint, and L. Hannah. 2013. Modeling plant species distributions under future climates: How fine scale do climate projections need to be? Global Change Biology 19:473–483.

Freire, J., C. Silva, S. Callahan, E. Santos, C. Scheidegger, and H. Vo. 2006. Managing rapidly evolving scientific workflows provenance and annotation of data. Pages 10–18 *in* L. Moreau and I. Foster, editors. International Provenance and Annotation Workshop 2006: Provenance and Annotation of Data. Springer, Berlin/Heidelberg, Germany.

Grechi, I., I. Chadès, Y. M. Buckley, M. H. Friedel, A. C. Grice, H. P. Possingham, R. D. van Klinken, and T. G. Martin. 2014. A decision framework for management of conflicting production and biodiversity goals for a commercially valuable invasive species. Agricultural Systems 125:1–11.

Hijmans, R. J. 2015. raster: geographic Data Analysis and Modeling. R package version 2.5-2. https://CRAN.R-project.org/package=raster

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

Hosmer, D. W., and S. Lemeshow. 2000. Applied logistic regression. Second edition. Wiley, New York, New York, USA.

Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46:10–18.

Hussey, M. A., and E. C. Bashaw. 1996. Performance of buffelgrass germplasm with improved winter survival. Agronomy Journal 88:944–946.

Ibarra-Flores, F. A., M. H. Martin-Rivera, J. R. Cox, T. A. Crowl, D. F. Post, R. W. Miller, and G. A. Rasmussen. 1995. Relationship between Buffelgrass survival, organic carbon, and soil color in Mexico. Soil Science Society of America Journal 59:1120–1125.

Jarnevich, C. S. 2018. Data for: forecasting an invasive species' distribution with global distribution data, local data, and physiological information: U.S. Geological Survey data release. https://doi.org/10.5066/p9y99uff

Jiménez-Valverde, A., A. T. Peterson, J. Soberón, J. M. Overton, P. Aragón, and J. M. Lobo. 2011. Use of niche models in invasive species risk assessments. Biological Invasions 13:2785–2797.

Marshall, V. M., M. M. Lewis, and B. Ostendorf. 2012. Buffel grass (*Cenchrus ciliaris*) as an invader and threat to biodiversity in arid environments: a review. Journal of Arid Environments 78:1–12.

Martin, T., H. Murphy, A. Liedloff, C. Thomas, I. Chadès, G. Cook, R. Fensham, J. McIvor, and R. van Klinken. 2015. Buffel grass and climate change: a framework for projecting invasive species distributions when data are scarce. Biological Invasions 17:3197–3210.

Martin-R, M. H., J. R. Cox, and F. Ibarra-F. 1995. Climatic effects on buffelgrass productivity in the Sonoran Desert. Journal of Range Management 48:60–63.

Morisette, J. T., C. S. Jarnevich, T. R. Holcombe, C. B. Talbert, D. Ignizio, M. K. Talbert, C. Silva, D. Koop, A. Swanson, and N. E. Young. 2013. VisTrails SAHM: visualization and workflow management for species habitat modeling. Ecography 36:129–135.

Olsson, A. D., J. Betancourt, M. P. McClaran, and S. E. Marsh. 2012. Sonoran Desert ecosystem transformation by a C4 grass without the grass/fire cycle. Diversity and Distributions 18:10–21.

Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37.

Pearce, J., and S. Ferrier. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modelling 128:127–147.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? Global Ecology and Biogeography 12:361–371.

Pearson, R. G., T. P. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27:285–298.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259.

Phillips, S. J., M. Dudik, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19:181–197.

Phillips, S., and J. Elith. 2010. POC plots: calibrating species distribution models with presence-only data. Ecology 91:2476–2484.

PRISM Climate Group. 2012. Oregon State University. http://www.prism.oregonstate.edu/

R Core Team. 2016. R: a language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria.

Rao, A., K. Singh, and J. Wight. 1996. Productivity of *Cenchrus ciliaris* in relation to rain-fall and fertilization. Journal of Range Management 49:143–146.

Schloerke, B., J. Crowley, D. Cook, F. Briatte, M. Marbach, E. Thoen, A. Elberg and J. Larmarange. 2017. GGally: extension to 'ggplot2'. R package version 1.3.1. https://CRAN.R-project.org/package=GGally

Seo, C., J. H. Thorne, L. Hannah, and W. Thuiller. 2009. Scale effects in species distribution models: implications for conservation planning under climate change. Biology Letters 5:39–43.

Stair, D., M. Dahmer, E. Bashaw, and M. Hussey. 1998. Freezing tolerance of selected Pennisetum species. International Journal of Plant Sciences 159:599–605.

Stevens, J., and D. Falk. 2009. Can buffelgrass invasions be controlled in the American Southwest? Using invasion ecology theory to understand buffelgrass success and develop comprehensive restoration and management. Ecological Restoration 27:417–427.

Thuiller, W., D. M. Richardson, P. Pysek, G. F. Midgley, G. O. Hughes, and M. Rouget. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11:2234–2250.

Wang, T., A. Hamann, D. Spittlehouse, and C. Carroll. 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. PLoS ONE 11:e0156720.

Ward, J. P. 2003. Estimating the potential distribution of buffelgrass in Saguaro National Park, Arizona: illustration of a conservation planning tool in the age of biotic homogenization. University of Arizona, Tuscon, Arizona, USA.

Warren, R. J., T. Ursell, A. D. Keiser, and M. A. Bradford. 2013. Habitat, dispersal and propagule pressure control exotic plant infilling within an invaded range. Ecosphere 4:art26.

Yesson, C., et al. 2007. How global is the global biodiversity information facility? PLoS ONE 2:e1124.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2279/full