# Predictive Mapping of Low-Density Juniper Stands in Prairie Landscapes of the Northern Great Plains

Authors: Kaskie, Kyle D., Wimberly, Michael C., and Bauman, Peter J.

Source: Rangeland Ecology and Management, 83(1) : 81-90

Published By: Society for Range Management

URL: https://doi.org/10.1016/j.rama.2022.03.005

Rangeland Ecology & Management 83 (2022) 81–90

Contents lists available at ScienceDirect

Rangeland Ecology & Management

journal homepage: www.elsevier.com/locate/rama

Original Research

# Predictive Mapping of Low-Density Juniper Stands in Prairie Landscapes of the Northern Great Plains

Kyle D. Kaskie[1], Michael C. Wimberly[2,*], Peter J. Bauman[1]

[1] Department of Natural Resource Management, South Dakota State University, Brookings, SD 57007, USA
[2] Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK 73019, USA

ARTICLE INFO

*Article history:*
Received 10 May 2021
Revised 19 February 2022
Accepted 21 March 2022

*Keywords:*
Eastern redcedar
Grassland ecology
Land cover change
Predictive vegetation mapping
Rocky Mountain juniper
Seedling establishment
Tree invasion

ABSTRACT

Expanding distributions of native juniper species have had significant ecological and economic impacts on prairie ecosystems of the Great Plains. Juniper encroachment reduces rangeland production by decreasing herbaceous biomass and affecting natural ecosystem functions as it alters other native plant communities, microclimates, and soils. Juniper distribution maps are needed to support proactive management, but they often underestimate the extent of low-density juniper stands. Our objectives were to extend a previous juniper mapping study by 1) fitting a predictive ecological model for low-density (< 15% fractional cover) juniper stands and assessing the classification accuracy, 2) determining the habitat variables that had the strongest associations with low-density juniper, and 3) applying the model to map low-density juniper stands, where proactive management has the greatest potential for stopping further juniper encroachment. The study area included counties bordering the Missouri River in southeastern South Dakota and northeastern Nebraska covering approximately 23 000 km$^2$. Environmental predictors included seed source distance and density, as well as topography, climate, soils, and land use variables. Areas of low-density juniper were identified by visual interpretation of sample plots from digital aerial photography. We used a machine-learning approach to classify low-density juniper with the random forests algorithm. Model accuracy was high with an area under the receiver operating characteristic curve of 0.884. Variables related to seed sources were the most important predictors, and precipitation, slope angle, and the local intensity of human land use also had substantial influences. A previous map based on Landsat imagery identified 209 968 acres (84 971 ha) as juniper with in the study area, and this study found an additional 430 648 acres (174 277 ha) classified as low-density juniper stands. These results can provide agencies and land managers with more accurate information about the distribution of juniper, and the underlying techniques can be extended to map woody plant encroachment in other areas.



## Introduction

In many parts of the North American Great Plains, expansion of woody plants threatens the natural functions of grassland ecosystems (Knapp et al., 2008; Van Auken 2009; Ratajczak et al., 2012). Notably, eastern redcedar *(Juniperus virginiana)* and Rocky Mountain juniper *(Juniperus scopulorum)* have affected carbon storage, soil characteristics, and plant communities within the prairie ecosystems of the Great Plains (Norris et al., 2001; McKinley and Blair 2008; Pierce and Reich 2010). In the central United States, the encroachment of these species (hereafter referred to collectively as "juniper") resulted in the conversion of 205 000 acres (82 960 ha) of non-forested land to forests between 2007 and 2012 (Meneguzzo and Liknes 2015). The shift from grassland to juniper cover affects ecosystem function and hydrology through increased soil infiltration and reduced streamflow (Zou et al., 2014; Zou et al., 2016). Carbon dynamics also change as carbon storage is shifted from belowground to aboveground (McKinley and Blair 2008). As juniper establishment increases, the microenvironments created by increasing tree cover affect the composition of understory plant communities and facilitate the encroachment of non-native species (Pierce and Reich 2010). These impacts of juniper encroachment have undesired consequences for grassland agriculture. Herbaceous productivity can be reduced up to 75%, which is comparable with forage levels at a heavily grazed site (Fuhlendorf et al., 2008). The economic and ecological consequences of juniper encroachment will continue to increase without efforts toward grassland conservation and restoration.

* Correspondence: Michael C. Wimberly, Dept of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK 73019, USA
*E-mail address:* mcwimberly@ou.edu (M.C. Wimberly).

https://doi.org/10.1016/j.rama.2022.03.005

Management of encroaching woody plants is crucial, especially when attempting to control a quickly expanding juniper footprint. Proactive and reactive management are two approaches for controlling juniper (Simonsen et al., 2015). Proactive management measures are implemented before juniper has established or is in vulnerable seedling and sapling stages and can include grazing, haying, and low-intensity prescribed burning (Wilson and Schmidt 1990; Smith 2011; Simonsen et al., 2015). In contrast, reactive management practices are implemented where juniper is already well established. Mechanical removal by timber cutting, herbicides, and high-intensity prescribed burning are commonly used methods for reducing juniper on established sites (Wilson and Schmidt 1990; Smith 2011; Simonsen et al., 2015). Reactive management of juniper is costly and time consuming, and it becomes less effective as stand density and tree size increase (Buehring et al., 1971; Ortmann et al., 1998; Bidwell et al., 2002). Therefore, proper planning is essential so that proactive management methods can be used in a timely manner.

Juniper maps can aid in targeting areas for proactive management, particularly if they can identify low-density sites before juniper is fully established. Newly established juniper can mature and begin producing seeds within 6 yr (Twidwell et al., 2021) and can reach 15% canopy cover within 10–15 yr (Fogarty et al., 2021). Remote sensing is a common way of obtaining spatial information about forest distribution and conditions that can be applied for management purposes (Franklin 2001; Giri 2012). Juniper distributions have been mapped with various remotely sensed data sources including very high spatial resolution (VHSR) aerial imagery (Anderson and Cobb 2004; Poznanovic et al., 2014), multispectral imagery (Sankey and Germino 2008; Kaskie et al., 2019), hyperspectral imagery (Wylie et al., 2000), and multisource fusion of data from active and passive sensors (Sankey et al., 2010; Wang et al., 2017). Although these studies have successfully mapped mature juniper stands, mapping areas with sparse coverage during the establishment phase has been more challenging. Wang et al., (2017) found that < 30% of juniper sites in Oklahoma with 10–20% tree cover was correctly classified using multitemporal synthetic aperture radar (SAR) and Landsat imagery. Kaskie et al., (2019) mapped junipers in South Dakota and Nebraska and similarly found that < 50% of juniper sites with 10–20% tree cover and < 15% with 1–10% cover were correctly classified using snow-covered winter Landsat imagery. Because of these challenges, tree cover estimates in rangelands have been largely absent from operational, continental-scale product such as the National Land Cover Dataset (Kaskie et al., 2019). More work is needed to improve the detection of areas in the early stage of juniper establishment where proactive management aimed at preventing juniper encroachment is most effective (Simberloff 2003; Yokomizo et al., 2009).

Several approaches have been used to map low-density tree cover in rangeland ecosystems. Techniques that use very-high-resolution imagery from digital aerial photographs (≤ 1-m pixel size) combined with object-based feature detection methods such as spatial wavelet analysis can detect individual trees (Strand et al., 2006; Falkowski et al., 2017). However, high data volumes and requirements for specialized software can make these methods challenging to apply. When used with machine learning techniques and large training datasets, Landsat can be used to predict fractional tree cover as a continuous variable (Allred et al., 2021). Landsat data are widely used for vegetation mapping because of their free availability, global coverage, and large historical archive. However, detecting low-density juniper stands with Landsat is difficult because the pixel size of Landsat (30 m) typically encompasses a mixture of juniper and other background materials. Alternatively, predictive vegetation mapping (Franklin 1995) can be used to supplement remotely sensed maps by identifying areas suitable for juniper establishment based on environmental variables. Predictive modeling of juniper encroachment in riparian habitats has found that juniper establishment is influenced by soils, floods, and the historical effects of flow regulation (Greene and Knox 2014; Illeperuma and Dixon 2021). However, riparian areas are only one of the many habitats where juniper occurs (Lawson 1990; Noble 1990), and larger-scale predictive models are needed for broader applications.

The goal of this study was to predict the probability of low-density juniper (< 15% cover) in areas of southeastern South Dakota and northeastern Nebraska that were not identified as containing juniper in a previous Landsat-based classification (Kaskie et al., 2019). This product reliably classified juniper pixels with fractional cover > 15% (89–95% overall accuracy) but had very low detection probabilities for sparser juniper cover. Thus, an additional modeling study focused on low-density juniper was undertaken to better characterize the extent of juniper encroachment across the study area. Specific objectives were to 1) fit a predictive ecological model for the probability of low-density juniper occurrence and assess the resulting classification accuracy, 2) determine habitat variables that have the strongest associations with low-density juniper stands, and 3) extrapolate predictions across the study area to identify areas where low-density juniper is most abundant.

## Methods

### *Study area*

Our study area covered 14 contiguous counties (nine counties in southeastern South Dakota and five counties in northeastern Nebraska [Fig. 1] all bordering the Missouri River. This area has a Köppen climate classification of *Dfa;* humid continental climate consisting of warm to hot summers and cold winters (Kottek et al., 2006), with an annual temperature range of 6–11°C and an annual average precipitation of 498–796 mm. Native vegetation consists of mixed-grass prairie species such as little bluestem *(Schizachyrium scoparium),* big bluestem *(Andropogon gerardii),* western wheatgrass *(Pascopyrum smithii),* sideoats grama *(Bouteloua curtipendula),* and green needlegrass *(Nassella viridula).* Woodlands are primarily found near drainages and riparian lowlands with the exception of small groves scattered across the prairie uplands. The most common deciduous species includes the plains cottonwood *(Populus deltoids)* with the occasional green ash *(Fraxinus pennsylvanica)* and American elm *(Ulmus americana).* Juniper species such as Rocky Mountain juniper *(Juniperus scopulorum)* and eastern redcedar *(Juniperus virginiana)* are also common (Barker and Whitman 1988). Steeply sloped drainages disrupt a flat to rolling topography composed largely of agriculture (48%) and herbaceous grasslands (39%), producing a fragmented landscape (Wimberly et al., 2018). Primary land uses include the agricultural production of corn, soybeans, and wheat, as well as cattle ranching (Wimberly et al., 2017).

### *Juniper training and validation data*

We used a stratified random sampling design to collect data from photo-interpreted juniper plots over a range of juniper densities. We allocated four strata, which included closed canopy woodlands, buffered closed canopy woodlands, planted shelterbelts (narrow rows of trees planted for wind protection), and nonwoodland areas. We digitized the closed canopy woodlands and planted shelterbelts in ArcGIS 10.5 following the guidelines outlined in Bauman et al., (2016) using 60-cm very high spatial resolution (VHSR) data from the National Agricultural Imagery Program collected in 2014 and 2016. To obtain samples of intermediate juniper densities, we

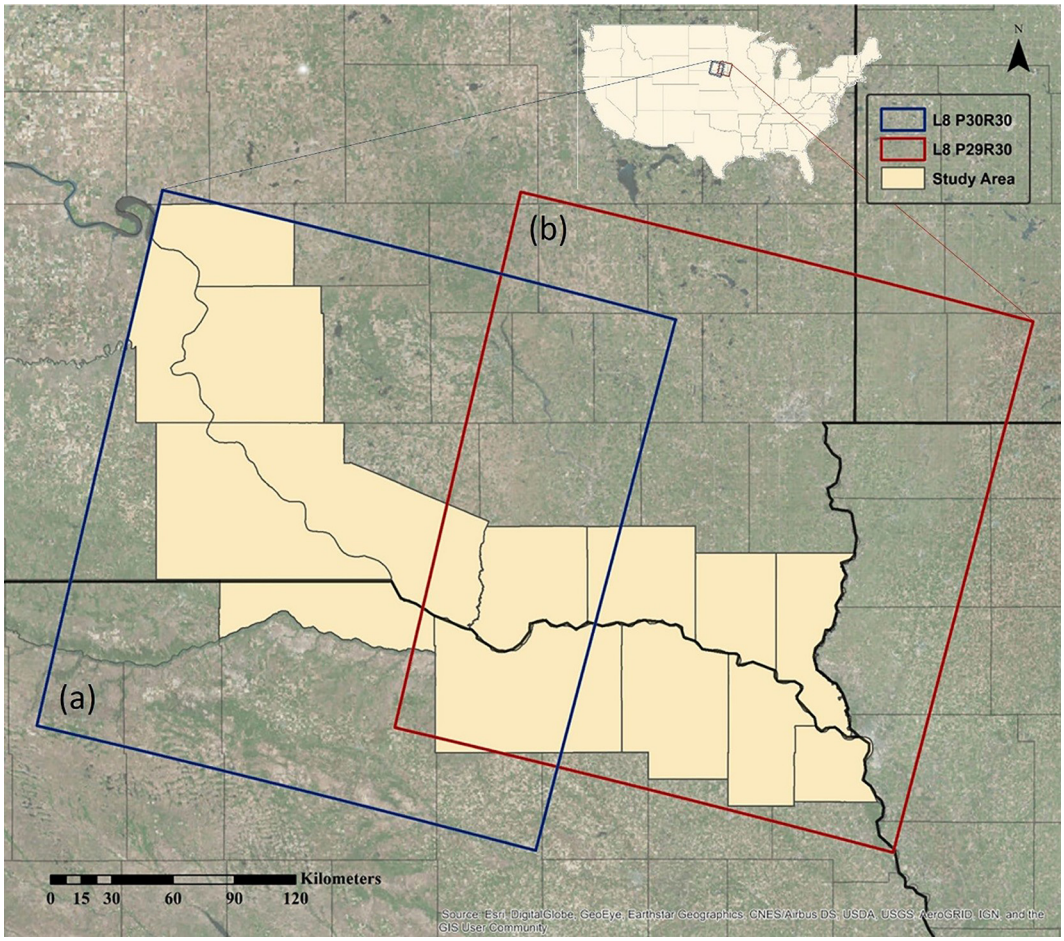


**Fig. 1.** Study area composed of nine counties in southeastern South Dakota and five counties in northeastern Nebraska. Landsat world reference system 2 path/rows: **a,** 30/30 and **b,** 29/30 cover the 14 contiguous counties.

placed a 90-m buffer around the digitized closed canopy stratum. The remaining nonwoodland areas had relatively few junipers.

We generated 4 308 random points including 1 282 points in closed-canopy woodlands, 671 points in shelterbelts, 1 210 points in buffered closed-canopy woodlands, and 1 145 points in nonwoodland areas. Each random point was referenced to a Landsat pixel by converting it to a $30 \times 30$ m polygon and snapping to the Landsat 8 pixel grid. We visually estimated juniper cover in each polygon using a combination of VHSR imagery, which included National Agricultural Imagery Program 2016 and other sources of winter imagery accessed through Google Earth from 2013 to 2017. All estimates were made by the same interpreter for consistency. These data were originally used for accuracy assessment of the remotely sensed juniper map developed by Kaskie et al., (2019). To increase the size of the dataset for training the juniper encroachment model, we sampled an additional 500 points in closed canopy woodlands and 500 points in buffered closed canopy woodlands. The final training dataset included all points with juniper cover $\leq$ 15% or no juniper cover. There were 865 low-density juniper points (1−15% cover) and 2 468 juniper absence points (Fig. 2).

*Ecological predictor variables*

We examined 15 predictor variables (Table 1) that were hypothesized to be associated with juniper encroachment. These variables measured topography (slope and aspect); climate (mean annual temperature and total annual precipitation); soils (percent sand, percent silt, percent clay, soil available water storage, depth to restricted layer, root zone depth, and soil drainage class); land use in the surrounding landscape (percent agriculture); seed sources (distance to juniper and percent juniper in the surrounding landscape); and flooding potential (distance to surface water). We processed these variables in ArcGIS 10.5 and resampled them all to a 30-m spatial resolution raster.

We used the National Elevation Dataset with 30-m spatial resolution to derive topographic indices. The National Elevation Dataset, created by the US Geological Survey, was accessed through the US Department of Agriculture Geospatial Data Gateway. Percent slope represented the physical gradient of the landscape and aspect indicated the direction the slope faced. We reclassified aspect into nine categories including flat (0% slope); cardinal directions (N, E, S, W); and ordinal directions (NE, SE, SW, NW).

We used the PRISM (Parameter-elevation Relationships on Independent Slopes Model) dataset to extract 30-yr normals (1981−2010) of climate variables, including average annual minimum temperature, average annual maximum temperature, and average annual precipitation. We used the minimum and maximum temperatures of each pixel to derive the mean annual temperature.

We used the gridded Soil Survey Geographic (gSSURGO) database obtained through the Geospatial Data Gateway to extract multiple soil characteristics. Gridded SSURGO contains the same data provided in the US Department of Agriculture Natural Resources Conservation Service SSURGO database in a 10-m spatial resolution raster format. Soil texture is represented as three sepa-

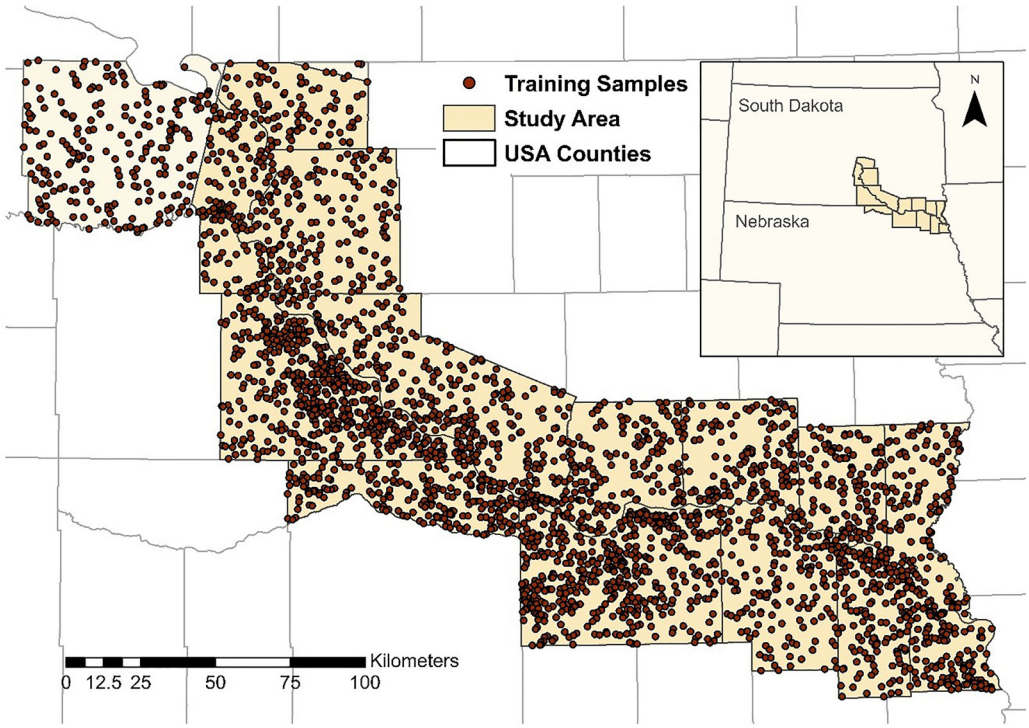


**Fig. 2.** Model training samples for random forests low-density juniper model.

**Table 1**
Predictor variables included in the random forests model of low-density juniper.

| Variables | Source | Description |
|---|---|---|
| Aspect | DEM | Direction a slope faces (1: Level; 2: North to 9: Northwest) |
| percSlope | DEM | Percent slope (rise divided by run multiplied by 100) |
| Precipann | PRISM | Total annual precipitation (in) |
| MAT | PRISM | Mean annual temperature (°F) |
| AWS0_150 | gSSURGO | Available water storage estimate from 0- to 150-cm depth; (cm) |
| Rootznemc | gSSURGO | Depth to which roots can extract soil water and nutrients |
| DRAINCLASS | gSSURGO | Drainage class based on frequency and duration of wet periods. |
| SAND | gSSURGO | Percent sand |
| SILT | gSSURGO | Percent silt |
| CLAY | gSSURGO | Percent clay |
| DEP2RESALYR | gSSURGO | Depth to restricted layer in soil profile |
| CDL_1_5 × 5 | CDL | Percent cultivated, alfalfa, or other hayfield pixels within a 5 × 5 window |
| DistJUNIP | Juniper Map | Euclidean distance to nearest juniper pixel (m) |
| JUNIP_5 × 5 | Juniper Map | Percent juniper pixels within a 5 × 5 window |
| DistWATER | NHD | Euclidean distance to nearest water body (m) |

CDL, Cropland Data Layer; DEM, Digital Elevation Model; gSSURGO; Gridded Soil Survey Geographic Dataset; NHD, National Hydrography Dataset; PRISM, Parameter-elevation Relationships on Independent Slopes Model.

rate factors including percent sand, percent silt, and percent clay (Wang et al., 2018). A combination of these three factors can be used to define the soil classification within the soil profile. We used the available water storage estimate, which represented the volume (mm) of plant available water the soil can store within a 0- to 150-cm soil profile. We used depth to restricted layer as a measure of the distance (cm) within the soil profile showing any restricting features that may constrain root growth or the movement of water and air. We used the root zone depth as a measure to the depth at which plants, root systems can effectively obtain nutrients and water. We also used soil drainage class, which is characterized into seven classes (excessively drained to very poorly drained) and reflects the natural frequency and duration of wet periods for the soil.

We obtained land cover data at a 30-m spatial resolution from the 2016 Cropland Data Layer (CDL). The CDL is produced annually and contains a remote-sensing based classification of different crop types. We used the 2016 CDL dataset to extract the percent agriculture within a 5 × 5-pixel window surrounding an individual 30-m pixel. We denoted agriculture as any human use of the landscape for cultivated crops, alfalfa, or other hay fields.

We used a 30-m classified juniper distribution map from a previous study (Kaskie et al., 2019) to estimate distance to seed sources. This map was based on two snow-covered Landsat 8 images that encompassed the study area. Juniper pixels were classified using a matched filtering approach in ENVI version 5.4 software (Exelis Visual Information Solutions, Boulder, CO). Matched filtering produces continuous values that indicate the similarity of each pixel to the spectral signature of pure juniper, and these values were thresholded to produce a classification of juniper presence/absence. The resulting map was validated using canopy cover data obtained from visual interpretation of very high resolution imagery. When juniper cover was above 50%, the detection probability was $\geq$ 90%. However, when juniper cover was lower than 20%, the true positive rate was $<$ 50%. We used this juniper presence-absence map to compute Euclidian distance to the near-

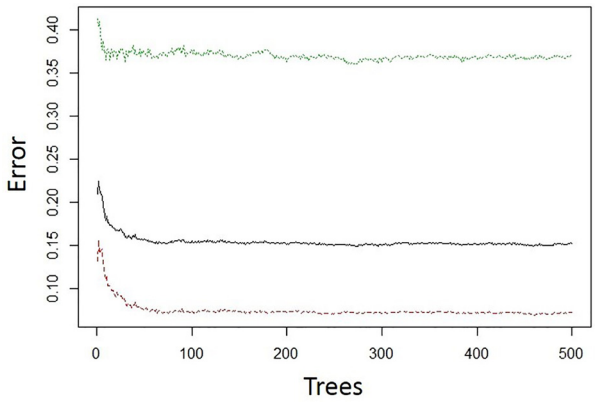


**Fig. 3.** Effects of number of trees (Ntree) on error rate for low-density juniper model (out-of-bag [OOB; *black line*]; juniper absence *[red line]*; and juniper presence *[green line]*).

est juniper pixel and the percent juniper within a 5 × 5-pixel window surrounding an individual 30-m pixel.

Distance to surface water was measured using the 1:24 000 scale US Geological Survey National Hydrography Dataset. We used the National Hydrography Dataset to map all surface water features within the study area. The Euclidean distance from each pixel to the closest surface water feature was then calculated.

### *Juniper encroachment model*

The random forests model is an ensemble-learning algorithm (Breiman 2001). This approach has been used in developing susceptibility models (Ismail et al., 2010; Wang et al., 2015; Youssef et al., 2016; Chen et al., 2017) and has been applied in numerous predictive mapping applications and across multiple disciplines (Prasad et al., 2006; Belgiu and Drăguţ 2016; Biau and Scornet 2016). We used random forests to predict the probability of low-density juniper presence given the set of observed environmental variables.

During each iteration of the random forests algorithm, the data are separated into two groups. The in-bag dataset consists of a bootstrap sample with approximately two thirds of the observations in the entire dataset and is used to train the algorithm. The in-bag samples are used to train a decision tree without pruning, where each split is based on one of a randomly selected set of predictor variables (Mtry). The out-of-bag (OOB) dataset contains the remaining observations and is used to generate predictions and estimate prediction error. Because the OOB predictions and associated error estimates are independent of the training dataset, they produce results that are similar to a leave-one-out cross-validation (Hastie et al., 2008). This process is repeated multiple times until a user-defined number of decision trees (Ntree) is reached. Each decision tree rule is used to cast a vote for a predicted class with the maximum number of votes becoming the final classification (Breiman 2001). The probability associated with a class is calculated as the total proportion of votes across all trees.

We developed random forests models using R statistical software version 3.4.1 (R Core Team 2021) and the randomForest package (Liaw and Wiener 2002). We set the Ntree parameter to 500, as trial runs with this value showed a stabilization of the errors before this maximum tree number was attained (Fig. 3). Increasing Ntree to 1000 or more had minimal effects on the results, so the parameter value of 500 was retained for computational efficiency. The Mtry parameter was set as the default, which is the square root of the total number of the predictor variables used in the model.

### *Model assessment*

The accuracy of model predictions was assessed by comparing observations with OOB predictions that were independent of the observations, providing results similar to a leave-one-out cross-validation (Hastie et al., 2008). We also computed accuracy statistics using a data-splitting approach, where 70% of points were used as training data and 30% were reserved as a validation dataset.

We used the Hosmer-Lemeshow goodness-of-fit test to evaluate model calibration. This test is commonly used in risk and susceptibility modeling (Bai et al., 2010; Catry et al., 2010; Fang et al., 2013). It categorizes subgroups (referred to as deciles of risk) and performs a Pearson chi-square statistic on the expected and observed frequencies. A close relation of the expected and observed frequencies reflects a model with good fit (Hosmer and Lemeshow 2000).

We evaluated model predictions using receiver operating characteristics (ROC). ROC examines multiple classification cutpoints by plotting the probability of detecting a true positive (sensitivity) against a false positive (1-specificity) over a range of cutoffs for classifying low-density juniper based on the proportion of votes from the random forests ensemble (Hosmer and Lemeshow 2000). The area under the curve (AUC) measures discrimination or the likelihood the model will predict target versus nontarget sites. Hosmer and Lemeshow (2000) suggest that AUC values represent discrimination as being none (0.5), acceptable (0.7–0.8), excellent (0.8–0.9), or outstanding (0.9–1.0).

We investigated the model performance of a full model and compared it with the performance of models with manually removed variables to eliminate unnecessary variables while improving our final model computation time (Plant 2012). We began by removing strongly correlated predictor variables followed by a trial-and-error process in which we independently excluded the remaining variables. We found the final model to have as much predictive power as our full model with a minimal effect on the modeling error.

## Results

### *Model accuracy*

The full model with all predictor variables achieved an out-of-bag (OOB) error rate of 15.3% (84.7% accuracy). After removing correlated and redundant variables, the final model contained eight predictor variables: distance to juniper, percent juniper in the surrounding landscape, total annual precipitation, percent slope, percent agriculture, soil available water storage, mean annual temperature, and aspect. The OOB estimate of error rate for the final model was 15.2% (84.8% accuracy), showing that the reduced model had as much predictive accuracy as our full model. The model had an OOB class error rate of 7.2% (92.8% accuracy) for nonjuniper (juniper absence) and 36.9% (63.1% accuracy) for low-density juniper (juniper presence).

We used the Hosmer-Lemeshow goodness-of-fit test to evaluate the calibration of our final model. The $P$ value for the Hosmer-Lemeshow test was $> 0.05$ ($\chi^2 = 5.8565$, df = 8, $P$ value = 0.6633), indicating that there was not a statistically significant difference between the expected and observed values and the final model was well calibrated. We also used the ROC to evaluate model predictions, and accuracy was high with an AUC of 0.884 (Fig. 4). The Hosmer-Lemeshow test and AUC results indicated that the final low-density juniper model was a good predictor of low-density juniper distributions.

Using the data-splitting validation approach, the overall error rate was 14.6% (85.4% accuracy). The class error rates were 6.3% (93.7% accuracy) for nonjuniper (juniper absence) and 38.6% (61.4%

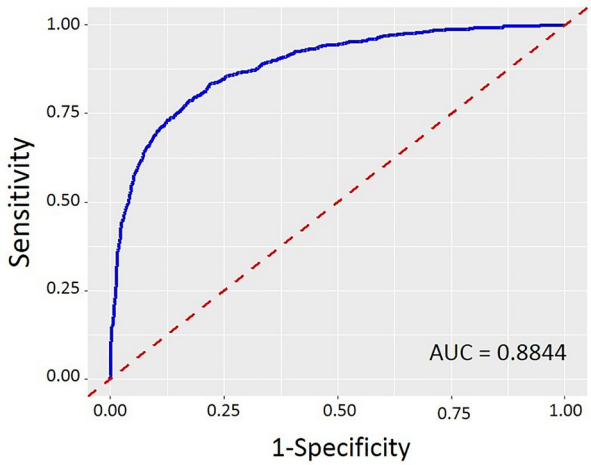


**Fig. 4.** Receiver operating characteristic curve to validate low-density juniper model.

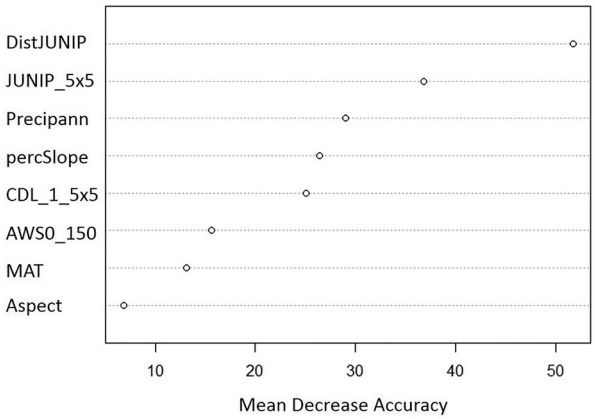


**Fig. 5.** Mean decrease accuracy of final eight low-density conditioning factors (listed in descending order) assigned by our random forests model. The eight predictor variables listed are distance to juniper, percent juniper in the surrounding landscape, total annual precipitation, percent slope, percent agriculture, soil available water storage, mean annual temperature, and aspect. Variable abbreviations are defined in Table 1.

accuracy) for low-density juniper (juniper presence). The AUC was 0.872. These results were similar to those based on the OOB predictions, confirming that accuracy remained high even when predicting observations that were outside of the training dataset.

*Environmental predictors*

The importance of the eight predictor variables is shown in Fig. 5 by descending order of mean decrease in accuracy. Distance to juniper was observed to have the highest conditional importance (51.71) with a strong negative relationship between low-density juniper and the distance to a seed source (Fig. 6a). The highest probabilities of low-density juniper (> 0.25) were within 100 m of a mature juniper seed source, and the probability decreased rapidly out to approximately 300 m. In the field, we observed that low-density junipers located longer distances (> 1 000 m) from a seed source were typically planted immature junipers. Percent juniper within a 5 × 5-pixel window (36.81) had the second highest importance with a strong positive relationship (see Fig. 6b), indicating that there was an increase in low-density juniper where multiple mature seed sources were in close proximity. Total annual precipitation (28.99) was observed to have a negative relationship (see Fig. 6c) and was followed by percent slope (26.40)

**Table 2**
Distribution of predicted probabilities for low-density juniper within the study area.

| Probability indices | Pixels | Acres | Hectares | Percent of total (%) |
|---|---|---|---|---|
| 0.000-0.075 | 15 386 464 | 3 421 871 | 1 384 782 | 62.6 |
| 0.075-0.176 | 4 937 870 | 1 098 156 | 444 408 | 20.1 |
| 0.176-0.380 | 2 308 263 | 513 344 | 207 743 | 9.4 |
| 0.380-0.694 | 1 157 495 | 257 422 | 104 175 | 4.7 |
| 0.694-1.000 | 778 916 | 173 226 | 70 102 | 3.2 |

with a positive relationship (see Fig. 5d) and percent agriculture (25.11) with a negative relationship (see Fig. 6e). The three variables with lowest importance included soil available water storage (15.67) with a negative relationship (see Fig. 6f), mean annual temperature (13.14) with a positive relationship (see Fig. 6g), and aspect with the lowest importance (6.86) and no clear relationship (see Fig. 6h).

*Juniper encroachment maps*

We produced a final low-density juniper map by grouping the continuous probability values into five Jenks natural breaks groups: 0.000–0.075, 0.075–0.176, 0.176–0.380, 0.380–0.694, and 0.694–1.000 (Fig. 7). By adjusting the probability thresholds for classifying low-density juniper, map users can control the tradeoff between detection probability (sensitivity) and correctly predicting juniper absence (specificity) as shown by the ROC curve in Fig. 4. For example, using probabilities > 0.38 (the two highest probability classes in Fig. 7) as a cutoff for low-density juniper prediction resulted in a sensitivity of 0.71 and a specificity of 0.88.

Visualizing the probabilities as a continuum provides information about the relative densities of low-density juniper pixels across the landscape. The probability map was able to capture areas of low-density, encroaching juniper that were not classified as juniper in the Kaskie et al., (2019) Landsat-based map (Fig. 8), with few low-density juniper pixels where probabilities were < 0.176 and increasing juniper density over probabilities ranging from 0.176 to 1. The summaries in Table 2 show that 430 648 acres (174 277 ha) had a predicted juniper probability > 0.38 in addition to the 209 968 acres (84 971 ha) of juniper originally classified in Kaskie et al., (2019). Thus, the footprint of juniper encroachment was considerably larger once the extent of low-density juniper stands was incorporated.

## Discussion

Satellite remote sensing has been shown to accurately predict the spatial distribution of relatively dense, mature juniper stands but is less effective at predicting locations where juniper is still establishing and cover is relatively low (Wang et al., 2017; Kaskie et al., 2019). In this study, we showed that it is possible to map low-density stands with juniper cover < 15%) using a model with predictor variables for topography, climate, soils, land use, and seed source availability. The model had a relatively high prediction accuracy (AUC = 0.88) and highlighted areas with low-density juniper that were not identified by our previous Landsat-derived map of juniper presence/absence. The predictive map of low-density juniper highlights locations where encroaching junipers are still primarily in seedling and sapling stages and are potentially suitable for proactive management, such as low-intensity burning, haying, and grazing, which can prevent the establishment of mature juniper that will be more difficult to remove.

Although the original Landsat-based map (Kaskie et al., 2019) was not effective at predicting low-density juniper stands, it allowed us to extract predictor variables that greatly influenced the juniper encroachment model. This was expected, as low-density ju-

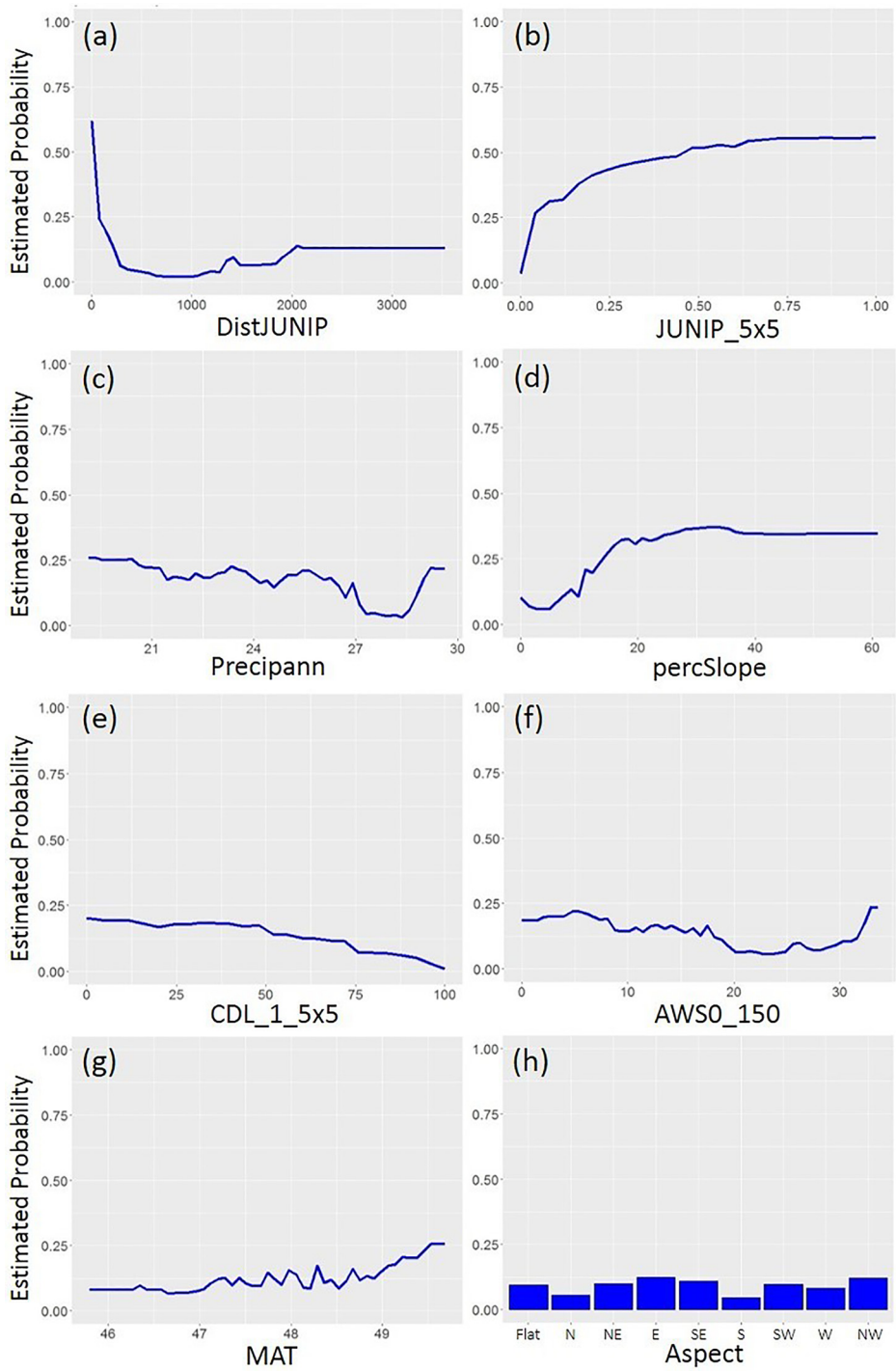


**Fig. 6. a-h,** Partial dependence plots for eight predictor variables used in the final random forests model for predicting low-density juniper. Variable abbreviations are defined in Table 1.

niper stands are usually associated with seed dispersal from established juniper sites (Holthuijzen et al., 1987; Yao et al., 1999). Holthuijzen and Sharik (1985) showed that within abandoned fields, eastern redcedar density decreased as the distance from a seed source increased. This result supports our finding that the probability of observing low-density juniper was highest < 100 m from established juniper and decreased with increasing distance from established juniper. Additionally, Owensby et al., (1973) found highest establishment of eastern redcedars within fields that were already heavily invaded by junipers. We similarly found that an increase in the distribution of established juniper within a $5 \times 5$-pixel window (≈60-m radius) increased the probability of low-density juniper occurrence. The fact that both variables were important in our model emphasizes that seed source density and distance to seed source are important predictors of juniper establishment.

Predictor variables related to juniper cover had the highest conditional importance within our model, but climatic and topographical factors were influential as well. The 30-yr normal of total annual precipitation had the third highest importance. We

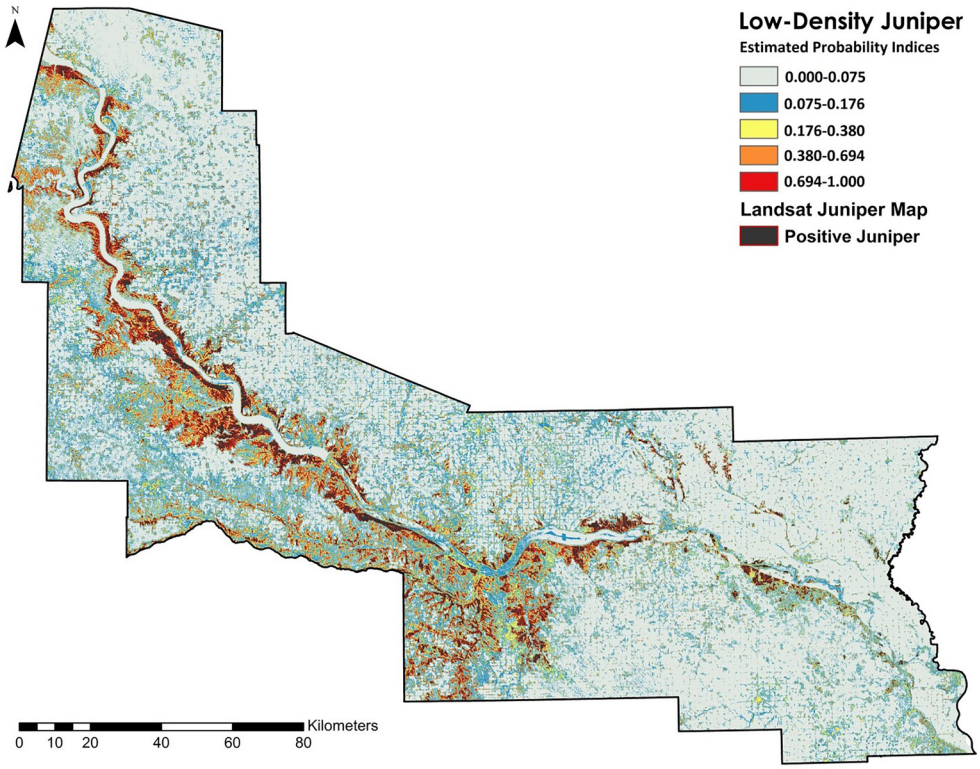


**Fig. 7.** Low-density juniper map covering the study area. Level of juniper probability is represented within five indices while including positive juniper classifications.

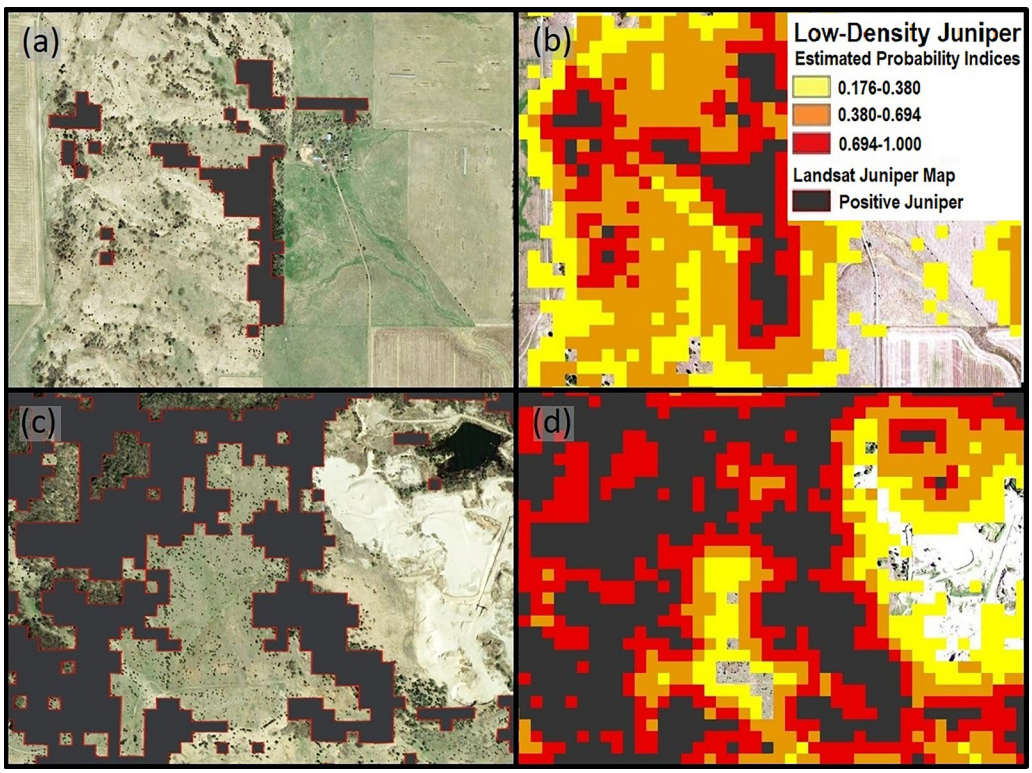


**Fig. 8.** A close-up view of the low-density juniper map. **a, c,** Original Landsat juniper classification. **b, d,** Probabilities of low-density juniper predicted by the model.

observed a slight negative effect for annual precipitation over a range of 18–27 in (457–686 mm). Owensby et al., (1973) similarly found that for every additional inch (2.54 cm) of precipitation, the juniper invasion rate decreased by 0.2 trees per acre (0.49 trees per ha). We also found a positive relationship with slope angle, in agreement with previous findings that Eastern redcedar can be associated with moderate to steep slopes (Anderson 2003), which have shallow soils and provide less competition from other plants and more protection from fire (Bryant 1989; Pierce and Reich 2010). The negative relationship with agricultural land use in a $5 \times 5$-pixel window suggests that activities such as active cultivation, haying, and grazing prevent juniper establishment and reduce the potential for spread into the surrounding landscape. Fire is another type of disturbance that can greatly reduce juniper abundance, particularly at the seedling and sapling stages. We did not have reliable maps of fire history for our study area, but where these data are available they could be incorporated as an additional predictor variable.

Variable importance ranking for our remaining variables fell considerably below the highest-ranked variables, yet removal resulted in a slight reduction in model performance and they were therefore retained in the final model. Low-density juniper occurrence decreased with increasing soil water storage but then increased again at water storage values $>$ 30%. Similar to the effects of slope, very low and high water storage values may indicate sites where competition from other plant species is minimized. Low-density juniper was also increased slightly at warmer temperatures, although there was only a small variation in mean annual temperature (46–50°F, 7.8–10°C) across the study area. Aspect was the least important environmental factor and did not have a clear relationship with low-density juniper. Some studies have shown aspect to influence juniper establishment (Lawson 1990; Schmidt and Stubbendieck 1993), whereas others have found no effects of aspect on juniper establishment (Tunnell et al., 2004).

### *Implications*

Previously developed juniper classification maps based on Landsat imagery allowed us to characterize current distributions of established juniper (Kaskie et al., 2019). The predictive model of low-density juniper developed in this study provided additional information about locations with that were not detectable with Landsat imagery. These low-density juniper stands were abundant within the study area. Adding the locations where low-density juniper was predicted with high probability ($>$ 0.38) to our previous Landsat-based map more than tripled the prediction of juniper extent from 84 971 to 259 248 ha. If these additional efforts were not made to map low-density juniper stands, the extent of juniper encroachment would be greatly underestimated.

The greater time and cost of managing juniper encroachment with increasing size and density emphasizes the importance of detecting and mapping these low-density areas (Ortmann et al., 1998; Bidwell et al., 2002). Emerging management frameworks for reducing woody plant encroachment are inherently spatial (Twidwell et al., 2021), and early detection and rapid response following woody plant recruitment has been identified as a core management philosophy. Predictive maps such as the one generated in this study can support these efforts by highlighting places in the landscape that contain low-density juniper abundance to prioritize them for proactive management. Estimates of the total area of low-density juniper at a particular location can be valuable for budgeting and determining equipment and personnel needed for management. Knowing the locations with low-density juniper can help to direct on-the-ground surveillance efforts and determine the management activities that will be most suitable at particular sites. Because seed sources are one of the most important constraints to juniper establishment, maps of the current juniper distribution are essential for understanding where dispersal and recruitment will occur in the future.

The techniques that we have used are distinctive from, but complementary to, other studies that have mapped rangeland tree cover as continuous variables (Falkowski et al., 2017; Allred et al., 2021) and detected rangeland regime shifts through spatial analysis of multitemporal land cover data (Uden et al., 2019). All the geospatial data that we used to develop the low-density juniper map are freely available for the entire United States and many other parts of the world. In addition, the modeling techniques used in the study have been implemented in a variety of freely accessible software platforms such as R (R Core Team 2021) and Google Earth Engine (Gorelick et al., 2017). Therefore, the approach that we developed and tested can be used by a wide variety of stakeholders, easily updated in the future, and extended to other rangeland systems where encroachment of juniper and other woody species are management concerns.

### *Research Data*

Geospatial datasets containing the Landsat-derived juniper map (https://doi.org/10.6084/m9.figshare.9241688.v1) and predicted low-density juniper probabilities (http://doi.org/10.6084/m9.figshare.17122316) are available from figshare. The results can be visualized via a Google Earth Engine app available at: (https://mcwimberly.users.earthengine.app/view/junipermaps).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

Allred, B.W., Bestelmeyer, B.T., Boyd, C.S., Brown, C., Davies, K.W., Duniway, M.C., Ellsworth, L.M., Erickson, T.A., Fuhlendorf, S.D., Griffiths, T.V., 2021. Improving Landsat predictions of rangeland fractional cover with multitask learning and uncertainty. Methods in Ecology and Evolution 12, 841–849.

Anderson, J.J., Cobb, N.S, 2004. Tree cover discrimination in panchromatic aerial imagery of pinyon-juniper woodlands. Photogrammetric Engineering and Remote Sensing 70, 1063–1068.

Anderson, M.D., 2003. *Juniperus virginiana*. Missoula, MT, USA: Fire Effects Information System. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory.

Bai, S.B., Wang, J., Lü, G.N., Zhou, P.G., Hou, S.S., Xu, S.N, 2010. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. Geomorphology 115, 23–31.

Barker, W.T., Whitman, W.C, 1988. Vegetation of the northern Great Plains. Rangelands 10, 266–272.

Bauman, P., Carlson, B., Butler, T., 2016. Quantifying undisturbed (native) lands in eastern South Dakota: 2013. South Dakota State University Extension, Brookings, SD, USA.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS Journal of Photogrammetric Remote Sensing 114, 24–31.

Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25, 197–227.

Bidwell, T.G., Weir, J.R., Engle, D.M, 2002. Eastern Redcedar control and management: best management practices to restore Oklahoma's ecosystems. Division of Agricultural Sciences and Natural Resources. Oklahoma State University, Stillwater, OK, USA.

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

Bryant, W.S., 1989. Redcedar (*Juniperus virginiana* L.) communities in the Kentucky River gorge area of the Bluegrass Region of Kentucky. USDA Forest Service North Central Forest Experiment Station General Technical Report NC-132, St. Paul, MN, USA. 304p.

Buehring, N., Santelmann, P.W., Elwell, H.M, 1971. Responses of Eastern Red Cedar to control procedures. Journal of Range Management 24, 378–382.

Catry, F.X., Rego, F.C., Bação, F.L., Moreira, F., 2010. Modeling and mapping wildfire ignition risk in Portugal. International Journal of Wildland Fire 18, 921–931.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z., Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151, 147–160.

Falkowski, M.J., Evans, J.S., Naugle, D.E., Hagen, C.A., Carleton, S.A., Maestas, J.D., Khalyani, A.H., Poznanovic, A.J., Lawrence, A.J, 2017. Mapping tree canopy cover in support of proactive prairie grouse conservation in western North America. Rangeland Ecology & Management 70, 15–24.
Fang, L.Q., Li, X.L., Liu, K., Li, Y.J., Yao, H.W., Liang, S., Yang, Y., Feng, Z.J., Gray, G.C., Cao, W.C, 2013. Mapping spread and risk of avian influenza A (H7N9) in China. Scientific Reports 3, 2722.
Fogarty, D.T., de Vries, C., Bielski, C., Dwidwell, T., 2021. Rapid re-encroachment by *Juniperus virginiana* after a single restoration treatment. Rangeland Ecology & Management 78, 112–116.
Franklin, J., 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progressive Physical Geography 19, 474–499.
Franklin, S.E., 2001. Remote sensing for sustainable forest management. CRC Press, Boca Raton, FL, USA, p. 424.
Fuhlendorf, S.D., Archer, S.A., Smeins, F., Engle, D.M., Taylor, C.A, 2008. The combined influence of grazing, fire, and herbaceous productivity on tree–grass interactions. In: van Auken, O. (Ed.), Western North American *Juniperus* communities. Springer, New York, NY, USA, pp. 219–238.
Giri, C.P., 2012. Remote sensing of land use and land cover: principles and applications. CRC Press, Boca Raton, FL, USA.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sensing and the Environment 202, 18–27.
Greene, S.L., Knox, J.C, 2014. Coupling legacy geomorphic surface facies to riparian vegetation: assessing red cedar invasion along the Missouri River downstream of Gavins Point Dam, South Dakota. Geomorphology 204, 277–286.
Hastie, T., Tibshirani, R., Friedman, J., 2008. The elements of statistical learning. Springer, New York, NY, USA, p. 745.
Holthuijzen, A.M., Sharik, T.L, 1985. The avian seed dispersal system of eastern red cedar *(Juniperus virginiana)*. Canadian Journal of Botany 63, 1508–1515.
Holthuijzen, A.M., Sharik, T.L., Fraser, J.D, 1987. Dispersal of eastern red cedar *(Juniperus virginiana)* into pastures: an overview. Canadian Journal of Botany 65, 1092–1095.
Hosmer, D.W., Lemeshow, S., 2000. Applied logistic regression, 2nd ed.. Wiley, New York, NY, USA, p. 375.
Illeperuma, N.D., Dixon, M.D, 2021. Analysing the influence of a large flood on eastern redcedar *(Juniperus virginiana)* distribution along the Missouri River using remote-sensing techniques. Ecohydrology e2279.
Ismail, R., Mutanga, O., Kumar, L., 2010. Modeling the potential distribution of pine forests susceptible to *Sirex noctilio* infestations in Mpumalanga, South Africa. Transactions in GIS 15, 709–726.
Kaskie, K.D., Wimberly, M.C., Bauman, P.J, 2019. Rapid assessment of juniper distribution in prairie landscapes of the northern Great Plains. International Journal of Applied Earth Observation 83, 101946.
Knapp, A.K., Briggs, J.M., Collins, S.L., Archer, S.R., Bret-Harte, M.S., Ewers, B.E., Peters, D.P., Young, D.R., Shaver, G.R., Pendall, E., Cleary, M.B, 2008. Shrub encroachment in North American grasslands: shifts in growth form dominance rapidly alters control of ecosystem carbon inputs. Global Changes in Biology 14, 615–623.
Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift 15, 259–263.
Lawson, E.R., 1990. *Juniperus virginiana* L. eastern redcedar. In: Burns, R.M., Honkola, B.H. (Eds.), Silvics of North America. Vol I: conifers. Agricultural Handbook 654. USDA Forest Service, Washington, DC, USA, pp. 131–140.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.
McKinley, D.C., Blair, J.M, 2008. Woody plant encroachment by *Juniperus virginiana* in a mesic native grassland promotes rapid carbon and nitrogen accrual. Ecosystems 11, 454–468.
Meneguzzo, D.M., Liknes, G.C, 2015. Status and trends of eastern redcedar *(Juniperus virginiana)* in the central United States: analyses and observations based on Forest Inventory and Analysis data. Journal of Forestry 113, 325–334.
Noble, D.L., 1990. *Juniperus scopulorum* Sarg. Rocky Mountain juniper. In: Burns, R.M., Honkola, B.H. (Eds.), Silvics of North America. Vol. I: conifers. Agricultural Handbook 654. USDA Forest Service, Washington, DC, USA, pp. 116–126.
Norris, M.D., Blair, J.M., Johnson, L.C, 2001. Land cover change in eastern Kansas: litter dynamics of closed-canopy eastern redcedar forests in tallgrass prairie. Canadian Journal of Botany 79, 214–222.
Ortmann, J., Stubbendieck, J., Masters, R.A., Pfeiffer, G.H., Bragg, T.B, 1998. Efficacy and costs of controlling eastern redcedar. Journal of Range Management 51, 158–163.
Owensby, C.E., Blan, K.R., Eaton, B.J., Russ, O.G, 1973. Evaluation of eastern redcedar Infestations in the northern Kansas Flint Hills. Journal of Range Management 26, 256–260.
Pierce, A.M., Reich, P.B, 2010. The effects of eastern red cedar *(Juniperus virginiana)* invasion and removal on a dry bluff prairie ecosystem. Biological Invasions 12, 241–252.
Plant, R.E., 2012. Spatial data analysis in ecology and agriculture using R. CRC Press, Boca-Raton, FL, USA, p. 648.
Poznanovic, A.J., Falkowski, M.J., Maclean, A.L., Smith, A., Evans, J.S, 2014. An accuracy assessment of tree detection algorithms in juniper woodlands. Photogrammetric Engineering and Remote Sensing 80, 627–637.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.
R Core Team, 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria Accessed 15 November, 2021..
Ratajczak, Z., Nippert, J.B., Collins, S.L., 2012. Woody encroachment decreases diversity across North American grasslands and savannas. Ecology 93, 697–703.
Sankey, T.T., Germino, M.J, 2008. Assessment of juniper encroachment with the use of satellite imagery and geospatial data. Rangeland Ecology & Management 61, 412–418.
Sankey, T.T., Glenn, N., Ehinger, S., Boehm, A., Hardegree, S., 2010. Characterizing western juniper expansion via a fusion of Landsat 5 Thematic Mapper and LIDAR data. Rangeland Ecology & Management 63, 514–523.
Schmidt, T.L., Stubbendieck, J., 1993. Factors influencing eastern redcedar seedling survival on rangeland. Journal of Range Management 46, 448–451.
Simberloff, D., 2003. How much information on population biology is needed to manage introduced species? Conservation Biology 17, 83–92.
Simonsen, V.L., Fleishmann, J.E., Whisenhunt, D., Volesky, J.D, Twidwell, D., 2015. Act now or pay later: evaluating the cost of reactive versus proactive eastern redcedar management. University of Nebraska Extension Institute of Agriculture and Natural Resources, Lincoln, NE, USA, p. 9.
Smith, S., 2011. Eastern red-cedar: positives, negatives and management. Samuel Roberts Noble Foundation, Ardmore, OK, USA.
Strand, E.K., Smith, A.M., Bunting, S.C., Vierling, L.A., Hann, D.B., Gessler, P.E, 2006. Wavelet estimation of plant spatial patterns in multitemporal aerial photography. International Journal of Remote Sensing 27, 2049–2054.
Tunnell, S.J., Stubbendieck, J., Huddle, J., Brollier, J., 2004. Seed dynamics of eastern redcedar in the mixed-grass prairie. Great Plains Research 14, 129–142.
Twidwell, D., Fogarty, D.T., Weir, J.R, 2021. Reducing woody encroachment in grasslands: a guide for understanding risk and vulnerability. Oklahoma Cooperative Extension Service, p. 32 Publication E-1054.
Van Auken, O.W, 2009. Causes and consequences of woody plant encroachment into western North American grasslands. Journal of Environmental Management 90, 2931–2942.
Uden, D.R., Twidwell, D., Allen, C.R., Jones, M.O., Naugle, D.E., Maestas, J.D., Allred, B.W, 2019. Spatial imaging and screening for regime shifts. Frontiers in Ecology & Evolution 7, 407.
Wang, J., Xiao, X., Qin, Y., Dong, J., Geissler, G., Zhang, G., Cejda, N., Alikhani, B., Doughty, R.B, 2017. Mapping the dynamics of eastern redcedar encroachment into grasslands during 1984–2010 through PALSAR and time series Landsat images. Remote Sensing and the Environment 190, 233–246.
Wang, J., Xiao, X., Qin, Y., Doughty, R.B., Dong, J., Zou, Z., 2018. Characterizing the encroachment of juniper forests into sub-humid and semi-arid prairies from 1984 to 2010 using PALSAR and Landsat data. Remote Sensing and the Environment 205, 166–179.
Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015. Flood hazard risk assessment model based on random forest. Journal of Hydrology 527, 1130–1141.
Wilson, J., Schmidt, T., 1990. Controlling eastern redcedar on rangelands and pastures. Rangelands 12, 156–158.
Wimberly, M.C., Janssen, L.L., Hennessy, D.A., Luri, M., Chowdhury, N.M., Feng, H., 2017. Cropland expansion and grassland loss in the eastern Dakotas: new insights from a farm-level survey. Land Use Policy 63, 160–173.
Wimberly, M.C., Narem, D.M., Bauman, P.J., Carlson, B.T., Ahlering, M.A, 2018. Grassland connectivity in fragmented agricultural landscapes of the north-central United States. Biological Conservation 217, 121–130.
Wylie, B.K., Meyer, D.J., Choate, M.J., Vierling, L., Kozak, P.K., Green, R.O, 2000. Mapping woody vegetation and eastern red cedar in the Nebraska Sand Hills using AVIRIS. AVIRIS Airborne Geoscience Workshop. JPL Publication 00-18.
Yao, J., Holt, R.D., Rich, P.M., Marshall, W.S, 1999. Woody plant colonization in an experimentally fragmented landscape. Ecography 22, 715–728.
Yokomizo, H., Possingham, H.P., Thomas, M.B., Buckley, Y.M, 2009. Managing the impact of invasive species: the value of knowing the density–impact curve. Ecological Applications 19, 376–386.
Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M, 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839–856.
Zou, C.B., Turton, D.J., Will, R.E., Engle, D.M., Fuhlendorf, S.D, 2014. Alteration of hydrological processes and streamflow with juniper *(Juniperus virginiana)* encroachment in a mesic grassland catchment. Hydrological Processes 28, 6173–6182.
Zou, C.B., Qiao, L., Wilcox, B.P, 2016. Woodland expansion in central Oklahoma will significantly reduce streamflows—a modelling analysis. Ecohydrology 9, 807–816.