



Research Paper

Cite this article: Russell JM et al. (2024) Modelling pinyon pine distribution in the northern Great Basin, USA. *Environmental Conservation* page 1 of 8. doi: [10.1017/S037689292400016X](https://doi.org/10.1017/S037689292400016X)

Received: 13 March 2024

Revised: 20 July 2024

Accepted: 20 July 2024

Keywords:

Gymnorhinus cyanocephalus; Idaho; machine learning algorithm; Nevada; Oregon; *Pinus edulis*; *Pinus monophylla*; pinyon jay; random forest; Utah


Corresponding author:

Jacob M Russell;

Email: jacobrussell15@outlook.com

© The Author(s), 2024. Published by Cambridge University Press on behalf of Foundation for Environmental Conservation. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Modelling pinyon pine distribution in the northern Great Basin, USA

Jacob M Russell , Robert A Miller, Jay D Carlisle and Gregory S Kaltenecker

Intermountain Bird Observatory, Boise State University, Boise, ID, USA

Summary

Pinyon–juniper woodlands are dry ecosystems defined by the presence of juniper (*Juniperus* spp.) and pinyon pine (*Pinus* spp.), which stretch over 400 000 km² across 10 US states. Certain areas have become unnaturally dense and have moved into former shrub and grasslands, while others have experienced widespread mortality. To properly manage these woodlands, sites must be evaluated individually and decisions made based on scientific information that is often not available. Many species utilize pinyon–juniper woodlands, including the pinyon jay (*Gymnorhinus cyanocephalus*), named for its mutualism with pinyon pine, whose population has declined by c. 2.2% per year from 1966 to 2022, an overall decrease of c. 71%. To increase the likelihood of further research progress, we propose a tool to model the distribution of pinyon pine at a finer scale than current woodland classification tools in the northern US Great Basin: a random forest model using geographical, ecological and climate variables. Our results achieved an accuracy of 93.94%, indicating high predictive power to identify locations of pinyon pine in north-eastern Nevada, the south-eastern corner of Oregon and southern Idaho. These findings can inform managers and planners researching pinyon pine, pinyon–juniper woodlands and potentially the pinyon jay.

Introduction

Pinyon–juniper woodland is a type of dry woodland ecosystem in North America (S Somershoe, personal communication 2024) defined by the presence of one or more species of juniper (*Juniperus* spp.) and pinyon pine (*Pinus* spp.; Nielsen 2009a). Present in 10 US states and covering over 400 000 km², pinyon–juniper woodland is the most common forest type in the American Southwest, yet it remains one of the least studied in North America (Tausch & Hood 2007, Nielsen 2009a, 2009b). There are concerns that long-term historical changes in climate, along with the impacts of Euro-American settlement, are changing pinyon–juniper woodland ecological dynamics, causing movement into former shrub and grasslands, creating unnaturally dense stands and decreasing tree establishment rates (Romme et al. 2009, RK Shriver et al., unpublished data 2024). It is unclear whether previous research, such as that on palaeoecology, dendroecology, ecophysiology, restoration and the effects of drought, has adequately captured the vast differences in the structure, composition and physical properties that occur across the pinyon–juniper woodland range (Hartsell et al. 2020). Therefore, widespread treatments might be ineffective in terms of efforts to create a diverse, healthy, productive and sustainable ecosystem, and sites must be evaluated individually, with decisions made based on scientific information about the woodlands, which is often not available (Kaufmann et al. 1994, Gottfried et al. 1995, Romme et al. 2009).

In the Great Basin and Colorado plateau, pinyon–juniper woodlands cover over 181 000 km² (Miller et al. 2019) and provide a valuable array of resources including firewood, pine nuts, livestock grazing, recreation and Indigenous peoples' ceremonial plant materials (Hartsell et al. 2020). The two most prominent species of pinyon pine in the pinyon–juniper woodlands are single-leaf pinyon (*Pinus monophylla*), which are red-brown and grow up to 14 m tall at elevations from 1000 to 2300 m on dry low-montane or foothills in the Great Basin and Mohave Desert borderlands, and the Colorado pinyon (*Pinus edulis*), which looks similar, is abundant in Colorado, New Mexico, Arizona and eastern Utah and grows up to 21 m tall on dry mountain slopes, mesas and plateaus at elevations from 1500 to 2100 m (Lanner 1981, Kral 1993).

Many plants and animal species utilize pinyon–juniper communities, including the pinyon jay (*Gymnorhinus cyanocephalus*), a medium-sized bird in the family Corvidae, named for its mutualism with pinyon pines (Balda & Marzluff 1992, Hartsell et al. 2020). With superb spatial memory (Stafford et al. 2006) and expandable throat pouches that fit 40–50 seeds (Johnson & Balda 2020), the pinyon jay serves as the primary long-distance disperser of pinyon pine seeds (Lanner 1981, Balda & Marzluff 1992, Somershoe et al. 2020). The bird's population is currently estimated at 770 000 (Somershoe et al. 2020), but numbers have decreased by c. 2.2% compounded per year from 1966 to 2022 (BBS 2022), an overall population loss of c. 71%, making it the fastest-declining pinyon–juniper bird species (Johnson & Sadoti 2023). The



pinyon jay has been listed as 'Vulnerable' (BirdLife International 2020) and was recently petitioned to be listed as endangered or threatened under the US Endangered Species Act (ESA) by Defenders of Wildlife (Estrella 2022).

The causes of these declines in populations are poorly understood (Somershoe *et al.* 2020) despite extensive knowledge of the species' breeding habits, social behaviour and spatial memory (Johnson & Balda 2020). Multiple causes of these declines have been hypothesized, including the past 70 years of pinyon–juniper land management (Johnson & Sadoti 2023), as well as the loss of the habitat due to climate change (Johnson & Balda 2020, Somershoe *et al.* 2020, Estrella 2022), around which most of the concern for the pinyon jay population revolves (Van Riper *et al.* 2014, Johnson & Balda 2020). Droughts and recent high temperatures have been associated with a decrease in pinyon pine cone production (Forcella 1981, Redmond *et al.* 2012, Parmenter *et al.* 2018, Wion *et al.* 2020), and climate-related impacts have also had a broader effect on pinyon–juniper woodlands (Breshears *et al.* 2005, 2009, Clifford *et al.* 2013, Meddens *et al.* 2015, Shriver *et al.* 2022), with predictions of increased mortality across the American south-west (Williams *et al.* 2010, McDowell *et al.* 2016), decreased growth rates (Williams *et al.* 2010) and changes to distribution patterns (Cole *et al.* 2008b). Some tree species in the pinyon–juniper woodlands will become increasingly vulnerable under future climate conditions, and tree populations are expected to decline due to changes in both recruitment and mortality rate, as well as other factors such as insect or other plant infestations (Noel *et al.* 2023).

Specific research is needed on the effects of pinyon–juniper woodland management, as well as on pinyon pine biology and woodland dynamics (Somershoe *et al.* 2020). In addition to increasing knowledge of pinyon pine ecology and management, a more accurate depiction of the distribution of pinyon pine would support further research into the potential threats facing the ESA-petitioned pinyon jay, as the lack of an accurate habitat model is a constraint to the species' conservation. While pinyon jays are of the highest conservation priority, many other animal species are also reliant upon pinyon–juniper woodlands (Evans 1988), and the benefits of research into pinyon–juniper ecology and distribution might be extendable to these other species associated with this habitat.

Currently, the best and most complete vegetation classification data covering the USA are those of Landfire (Landscape Fire and Resource Management Planning Tools), a shared wildland fire management programme between the US Department of Agriculture Forest Service and the US Department of the Interior. However, Landfire is not specifically designed for or tuned to our study area, and there are instances in its Existing Vegetation Type (EVT) data that contain classifications for pinyon–juniper and juniper-only woodlands that are not ecologically accurate (NatureServe 2017, Landfire 2022). As an example, the pure juniper in the South Hills of southern Idaho, where there is no known instance of pinyon pine, is listed as pinyon–juniper woodland in Landfire. An opportunity exists, therefore, to build upon Landfire's classification power and improve its resolution and accuracy for locations of pinyon pine in pinyon–juniper woodlands.

We aimed to use a machine learning technique with geographical, ecological and climate variables to create a probability distribution of pinyon pine throughout the northern Great Basin, USA. We used this smaller region of single-leaf and Colorado pinyon pines' full range, as the model will be more

accurate when focused on an area that shares some biophysical similarities, and using a broader model might risk losing some specificity and accuracy. In addition, this region approximately represents the scope of the Intermountain Bird Observatory's work and where we have ecological knowledge of and relationships with forest services and other field experts. This prediction will provide future research efforts with a tool to pinpoint the areas of highest likelihood for pinyon pine in the broader pinyon–juniper landscape and could potentially benefit pinyon jays by supporting habitat modelling, improving pre-survey planning and refining upcoming occurrence models (S Somershoe, personal communication 2024). Understanding pinyon pine distribution will increase the knowledge surrounding pinyon–juniper woodlands and the threats they face, and it will ideally reduce the need for ground truthing (S Somershoe, personal communication 2024) and lead to the more rapid implementation of management and conservation actions.

Methods

Study area

Our study area was 365 300 km² of the northern Great Basin, including the north-western and north-eastern areas of Utah and Nevada, respectively, south and central Idaho and eastern Oregon (Fig. 1).

Acquisition of spatial data

We downloaded data from the Landsat 8 satellite data archive for Surface Reflection (SR) bands 2–7 (blue, green, red, near-infrared, short-wave infrared 1 and short-wave infrared 2) from the United States Geological Survey (USGS) LandsatLook tool (<https://landsatlook.usgs.gov/explore>). The images for each of the bands were captured between 1 June and 30 September 2022 and with less than 5% cloud cover. We acquired a digital elevation model (DEM) for our study area using The National Map (TNM) tool of the USGS at 1 arcsecond for 2022 (<https://apps.nationalmap.gov/downloader/>). We obtained EVT classification data from the Landfire programme 2020 version with 30 m × 30 m resolution (<https://landfire.gov/evt.php>). We downloaded 11 climate variables from the set of 30-year (1991–2020) Normals (precipitation, mean temperature, maximum temperature, minimum temperature, mean dew point temperature, maximum vapour pressure deficit, minimum vapour pressure deficit, solar radiation (horizontal, sloped and clear sky) and cloud transmittance) from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) Climate Group (<https://prism.oregonstate.edu/normals/>). We also downloaded long term (2081–2100) CMIP6 future climate projection data for mean, maximum and minimum temperatures from WorldClim (https://www.worldclim.org/data/cmip6/cmip6_clim30s.html) for the Shared Socio-Economic Pathway (SSP) 5-8.5 climate scenario.

We loaded these datasets as raster layers using QGIS 3.30.2 (QGIS Development Team 2023), re-projected to the EPSG-4269 coordinate referencing system using the NAD83 datum and cropped to the extent of the study area. All numeric datasets (e.g., climate data) were re-projected and aligned with the DEM using bilinear interpolation, whereas all categorical datasets (e.g., Landfire EVTs) were re-projected and aligned using a nearest neighbour approach.

We evaluated 22 predictor variables in our modelling that we hypothesized might influence pinyon pine presence (Table 1). We

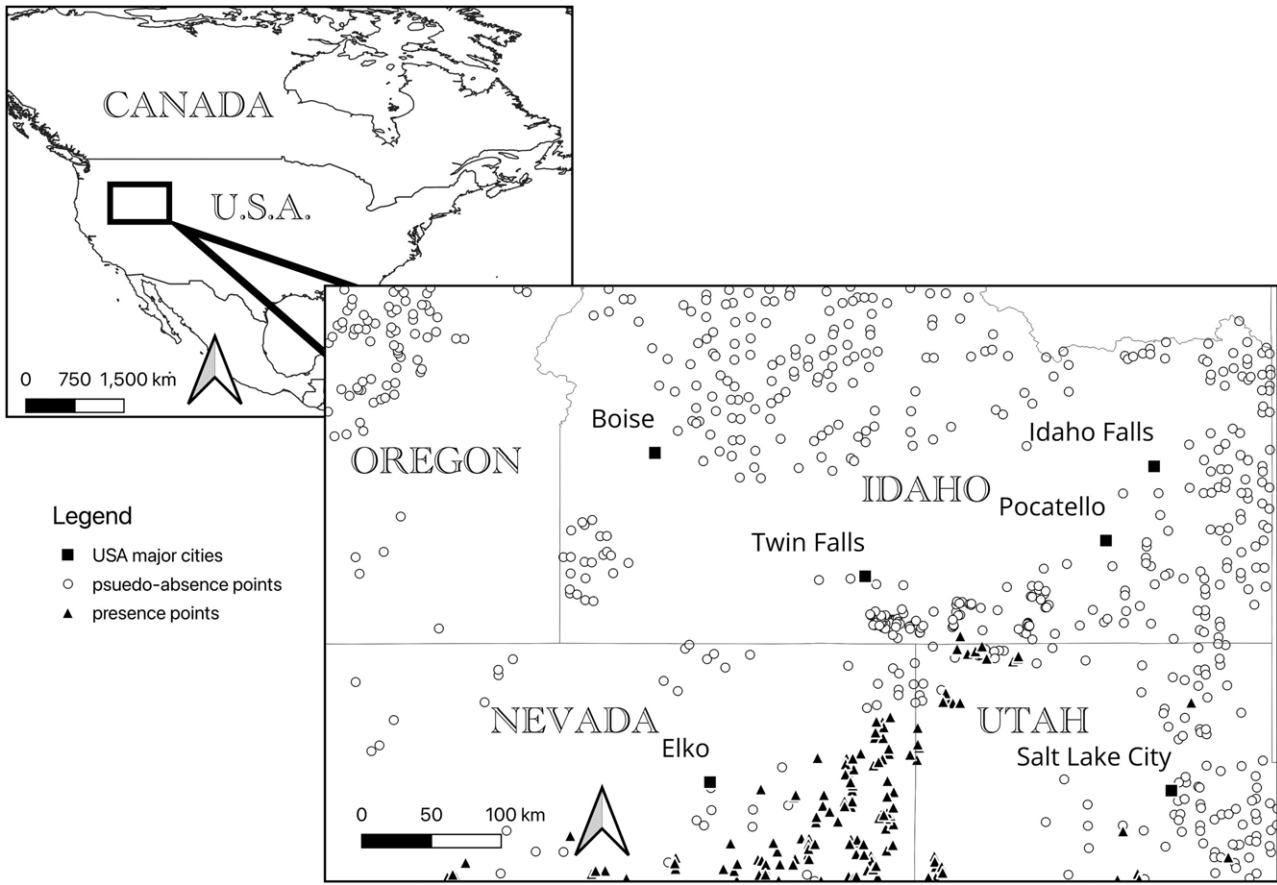


Figure 1. Pinyon pine study area within four western states of the USA with presence and pseudo-absence points and major cities. Extent (-119.00, 40.00 : -111.00, 45.00).

included the elevation and terrain variables (elevation, slope, terrain position index (TPI) and roughness), as single-leaf and Colorado pinyon pines are both found on varying slopes (Meeuwig 1979), topographical positions (Miller et al. 1981), mountainous terrain (Rust 1999) and in the elevation range of 1000–3000 m (Meeuwig 1979, Miller et al. 1981, Burns & Honkala 1990, West et al. 1998). We included the climate variables as pinyon pines are known to favour arid and semi-arid environments (Montes et al. 2019) and the current distributions of single-leaf and Colorado pinyon pines are limited by climate variables (Cole et al. 2008a). Although the Landfire EVT dataset is not sufficient to classify pinyon pine based on the granularity of our research, we included it because it distinguishes pinyon–juniper woodland from other forest types. Finally, we included surface reflectance bands because pinyon pine has a slightly different colour from other trees and the juniper that it often grows with.

Acquisition of field data

Field experts from the US Forest Service and the City of Rocks National Preserve located in south-central Idaho provided a dataset of pinyon pine presence and pseudo-absence points. The presence points were known instances of pinyon pines being observed and recorded, while the pseudo-absence points consisted of both verified locations as well as randomly generated locations. The randomly generated locations were selected from areas where the experts had spent many years conducting field research and not once seen an instance of pinyon pine, so we have an extremely high level of confidence in them and consider them much closer to true

absences. The dataset contained 144 presence points and 621 pseudo-absence points. During analysis, we included all presence points and randomly selected 288 of the 621 pseudo-absence points for use to maintain a ratio of 1:2 to reduce modelling bias (Evans et al. 2011).

Random forest modelling

We modelled pinyon pine presence using random forest classification modelling with the package ‘randomForest’ (Breiman 2001) in R version 4.3.0 (R Core Team 2023). In randomForest, many independent but identically distributed decision trees first cast a vote for a class, and then the most popular class becomes the model’s prediction (Breiman 2001, Yiu 2019). Ten parallel models were run, so predictions were made on 2880 pseudo-absence and 1440 presence points.

To identify which geographical, ecological and climate variables were predictive of pinyon pine presence, we combined satellite imagery and field data into a single dataset (Table 1). We found that all of the predictor variables were important to the model using the variable selection algorithm in the R package ‘rfUtilities’ (Murphy et al. 2010, Evans et al. 2011, Behl & Benkman 2018), and therefore we included all of them in the final model.

We mapped predicted probability of habitat suitability for pinyon pine across our study area for both current and future climate conditions by replacing three variables in our model (mean, maximum and minimum annual temperature) with their future estimates. The future projection made critical assumptions – such as no change to current land cover or other variables due to

Table 1. Predictor variables, type, units, range, mean, standard deviation (SD) and definition used in modelling of pinyon pine presence within the northern Great Basin, USA.

Name	Variable type	Units	Range	Mean	SD	Description
Elev	Terrain	Metres	1102.73–3158.13	2024.35	338.86	Elevation above mean sea level
Slope	Terrain	Degrees	0.02–43.33	15.32	9.44	Calculated using four surrounding points (Fleming & Hoffer 1979)
TPI	Terrain	NA	–5.83 to 4.49	–0.03	1.28	Terrain position index relative to the surrounding 150 m (Wilson et al. 2007)
Roughness	Terrain	Metres	0.02–67.79	19.50	12.81	Difference in elevation minimum and maximum within 150 m (Wilson et al. 2007)
Precip	Climate	Millimetres	237.48–1532.82	621.50	229.36	Average annual precipitation (rain + melted snow)
MinTemp	Climate	°C	–5.85 to 5.17	–0.61	2.03	Average annual minimum temperature
MaxTemp	Climate	°C	6.16–18.09	12.40	2.28	Average annual maximum temperature
MeanTemp	Climate	°C	–0.21 to 11.63	5.89	2.05	Average annual mean temperature
TDMean	Climate	°C	–9.51 to 0.68	–3.98	1.55	Average annual mean dewpoint temperature
VPDMin	Climate	Millibars	0.43–4.58	2.33	0.77	Average annual minimum vapour pressure deficit
VPDMax	Climate	Millibars	7.34–19.72	12.86	2.49	Average annual maximum vapour pressure deficit
SolClear	Climate	MJ/m ² /day	19.38–22.52	21.19	0.60	Average daily solar radiation received on horizontal surface on clear days
SolSlope	Climate	MJ/m ² /day	13.01–17.67	15.73	0.91	Average daily solar radiation received on sloped surface on all days
SolTotal	Climate	MJ/m ² /day	13.99–17.70	15.84	0.88	Average daily solar radiation received on horizontal surface on all days
SolTrans	Climate	Ratio	0.66–0.80	0.73	0.03	Average transmittance (cloudiness) (SolTotal/SolClear)
Landfire	Vegetation	Categorical	NA	NA	NA	Landfire Existing Vegetation Type
Band2	Imagery	Micrometres	7197.07–13 394.22	8755.96	1046.61	Surface Reflection, Band 2 – Blue
Band3	Imagery	Micrometres	7493.25–15 942.38	9598.15	1368.51	Surface Reflection, Band 3 – Green
Band4	Imagery	Micrometres	7435.51–18 508.74	10 045.47	1834.55	Surface Reflection, Band 4 – Red
Band5	Imagery	Micrometres	7438.06–25 012.37	15 422.81	3117.30	Surface Reflection, Band 5 – Near Infrared
Band6	Imagery	Micrometres	7394.04–24 412.57	13 582.82	3362.176	Surface Reflection, Band 6 – Short-Wave Infrared 1
Band7	Imagery	Micrometres	7374.27–20 300.54	11 700.88	2764.94	Surface Reflection, Band 7 – Short-Wave Infrared 2

NA = not applicable.

climate change – that we understand are unlikely but are necessary to create a conservative estimate of the potential effect of a changing climate on pinyon pine habitat suitability. We arbitrarily selected 0.66 as a benchmark and compared overall area of the current prediction to the future climate projection prediction and to the area Landfire has classified as pinyon–juniper woodland.

Measures of model performance

We computed a confusion matrix (with true and false positives, negatives and overall accuracy) to evaluate the performance of the random forest model. Overall accuracy was calculated by dividing the sum of correctly classified locations by the total number of locations. It is important to note that this is a theoretical measure and does not represent ecological accuracy. We also plotted the receiver operating characteristic (ROC) curve to calculate the value of the area under the curve (AUC). The ROC curve illustrates the ratio, at any given point, between the model's true- and false-positive rates (Florkowski 2008, Hajian-Talaki 2013). AUC values of 0.5 suggest no ability to discriminate between positives and negatives, values of 0.7–0.8 indicate acceptable levels of discrimination, values of 0.8–0.9 are considered excellent and values above 0.9 are considered outstanding (Mandrekar 2010).

We calculated the model's true skill statistic (TSS), which compares the number of correct predictions minus those attributable to random guessing, and Cohen's kappa coefficient, which also corrects the overall accuracy of the model using the accuracy expected by chance (Allouche et al. 2006). Both statistics range from –1 to +1, with +1 indicating perfect agreement and values less than or equal to 0 indicating random performance at best (Cohen 1960, Allouche et al. 2006).

Results

Using geographical, ecological and climate data, the random forest algorithm showed great potential for predicting the distribution of pinyon pine throughout the northern Great Basin, USA. The variables ranked highest in importance, as explained by mean decrease in accuracy (MDA; Han et al. 2016), included the four solar radiation variables, climate variables and elevation (Fig. 2). The solar transmittance (SolTrans) and radiation (SolSlope and SolTotal) variables all indicate increased likelihood of presence with higher values (Fig. 3). Lower values of precipitation showed higher likelihood for habitat suitability for pinyon pine. Probability was highest for elevations of 1500–2500 m, peaking at c. 2000 m. Maximum likelihood occurred at an average mean dewpoint temperature of c. –5°C (Fig. 3).

We generated a study-wide prediction for the probability of habitat suitability for pinyon pine occurrence, with the highest-probability areas including the mountain ranges of northern Nevada and north-western Utah (Fig. 4). When selecting only predictions greater than or equal to 0.66 (an arbitrary choice), 21 006.6 km² of the study area indicated habitat suitability, but only 4169.79 km² of that was classified as pinyon–juniper woodland by the Landfire EVT data. In addition, of the 8548.35 km² of Landfire-classified pinyon–juniper woodlands, 3406.13 km² were not shared by the predictions. That is not to say that Landfire has misclassified nearly 3500 km², nor that it has missed over 16 800 km² of pinyon–juniper woodland. Rather, those areas are locations that have been classified as pinyon–juniper woodland in Landfire but have low probability for habitat suitability and as those that Landfire has not classified as pinyon–juniper woodland but have high probability for habitat suitability.

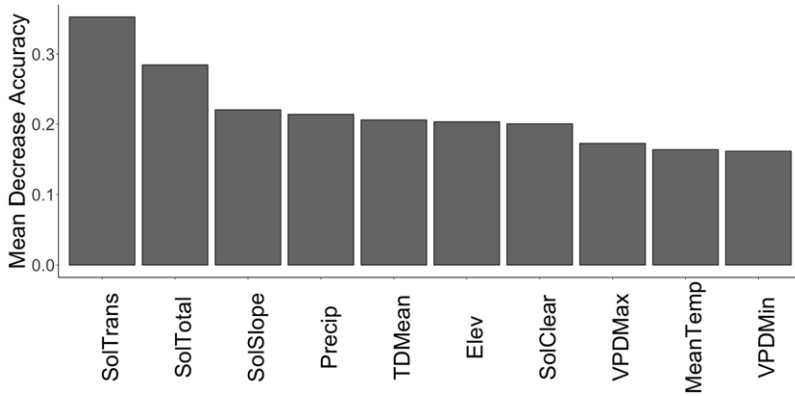


Figure 2. The relative importance of the 10 most important variables used in predicting pinyon pine presence within the northern Great Basin, USA, as measured by random forest classifiers using the mean decrease in accuracy. (See Table 1 for variable definitions.)

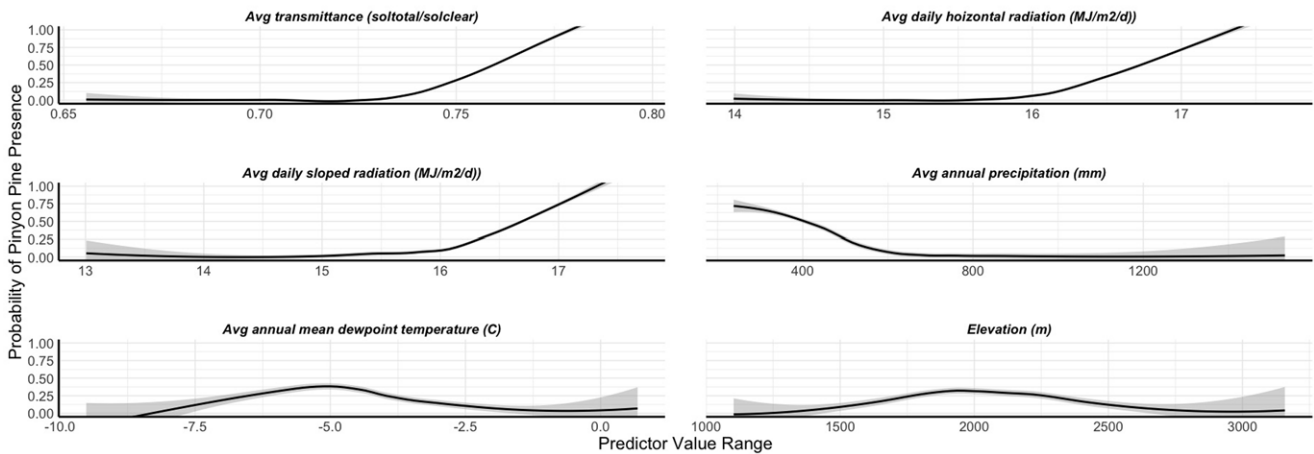


Figure 3. Marginal effects plots for the six most important variables used in predicting pinyon pine presence within the northern Great Basin, USA. (See Table 1 for variable definitions.)

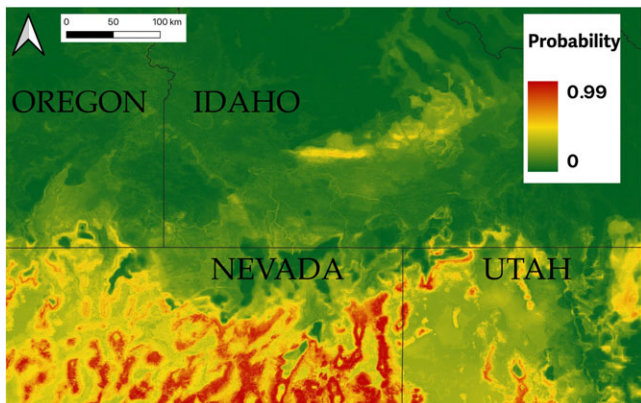


Figure 4. Map of the predicted probability of presence of pinyon pine within the northern Great Basin, USA.

Upon replacing just three climate variables in the model (maximum, minimum and mean annual temperature) with future projections of those variables under an SSP5-8.5 climate model dataset, we found that the area of habitat with high probability of suitability for pinyon pine was projected to significantly decrease. The overall area of predictions greater than or equal to 0.66 (again

an arbitrary selection) decreased from 21 006.6 to 10 438.44 km², not shifting greatly in geographical range, but shrinking by 50.31% in the current areas of highest prediction in the hills of northern Nevada and north-western Utah.

Model performance

The average overall accuracy of the random forest model was 93.94%, with a false-positive rate of 0.05 and a false-negative rate of 0.09 (Table 2). The prediction counts (2880 for pseudo-absence and 1440 for presence) were 10 times the sample size (288 pseudo-absence and 144 presence) as we ran 10 parallel models. The AUC value was 0.96, classifying our model as outstanding, and the TSS and Cohen’s kappa coefficient were 0.87 and 0.88, respectively.

Discussion

Our objective in this work was to create a probability distribution of pinyon pine throughout the northern Great Basin, USA, by using a machine learning technique with geographical, ecological and climate variables. With these results showing a model with high accuracy and strong classification capability, we were very successful in achieving that objective. This prediction map will assist further research efforts in pinpointing probable areas of

Table 2. Confusion matrix generated from random forest model predicting presence of pinyon pine within the northern Great Basin, USA.

	Predicted absence	Predicted presence	Error rate	Interpretation
Pseudo-absence	2742	138	0.05	False-positive rate
Presence	124	1316	0.09	False-negative rate
				Overall accuracy = 93.94%

pinyon pine in the broader pinyon–juniper woodlands, help increase our knowledge of pinyon–juniper woodlands and could prove useful for future pinyon jay research.

The model was trained on the realized niche of pinyon pine, not the fundamental niche, so the results could be stemming from direct environmental conditions, unquantified biological interactions or some combination of both. The top variables from the model were all climate variables, specifically solar radiation, which indicates a narrow climate envelope, but could be the result of pinyon pine being outcompeted by other species in higher-elevation, more moist environments and being unable to successfully maintain populations in more arid environments, for example, and there may have been other important variables that we did not include in our model. Another study found that certain soil compositions and other climate variables (e.g., mean temperatures of the wettest and coldest quarters and precipitation of the driest month) were among the most important variables when predicting habitat suitability for both single-leaf and Colorado pinyon pines across their entire range under climate change (Buck & Flores-Rentería 2022).

For the top three variables in our model (SolTrans, SolTotal and SolSlope), as well as the precipitation variable (Precip), the increasing presence likelihood with increasing solar radiation and decreasing precipitation aligns with pinyon pine being a desert pine species that prefers arid environments (Lanner 1981, Kral 1993, Montes *et al.* 2019). The elevation (Elev) variable was also important, indicating higher likelihood in the known elevation range (1000–2300 m) of pinyon pine (Lanner 1981, Kral 1993), but it is not surprising that it has less importance than others, as the same conditions appear at different elevations as latitude changes. We did expect the EVT variable (Landfire) to rank higher, as it is known to be suitable for distinguishing between types of woodlands, but there may be some confounding issues (e.g., large areas of known juniper classified as pinyon–juniper woodland). Altogether, however, the results of our model indicate higher probabilities of habitat suitability in areas with lower levels of precipitation and higher levels of solar radiation, which is to be expected in pinyon pine.

Models created with few presence points for large study areas run the risk of being biased and therefore not capturing all of the relevant information necessary to accurately predict across a wide range of conditions for the target species (Stockwell & Peterson 2002, McPherson *et al.* 2004, Hernandez *et al.* 2006). However, the random forest method has been shown to have the best performance when compared to other common machine learning algorithms, a better fit to testing data and to provide better species' range maps than other commonly used methods, and it has been known to perform extremely well in terms of ecological predictions, making it suitable for applications to conservation, especially when working with an under-sampled study area (Mi *et al.* 2017). Increased study efforts focused on pinyon pine and pinyon–juniper woodlands will naturally increase the number of presence points available, which can then be used to refine the working model.

Although external cross-validation is often performed to supplement the metrics of a random forest model, we used only the built-in performance evaluation. The high accuracy of the random forest model does not necessarily translate to similarly accurate ecological predictions, but it is an indication of a model that might prove helpful. It is therefore important to consider the consequences of false positives (travelling to a location where there is no pinyon pine) and false negatives (refraining from travelling to a location that does have pinyon pine). The latter has a worse effect, as the goal of further research is to find and study the largest amount of pinyon pine and pinyon–juniper woodland possible. These errors could also have downstream effects on analyses using these (or similar) data to study species that utilize pinyon–juniper woodlands, as false positives can overestimate available resources and false negatives can underestimate them. However, since the model performance was excellent and both the false-positive and false-negative rates were low, we expect these incorrect predictions to be manageable at the study area level, and we suggest that this model would greatly reduce the time needed for ground truthing of study sites in this region.

We believe this modelling effort can also augment data collection relevant to the proposed listing of the pinyon jay as threatened or endangered under the ESA, encouraging the prioritization of management actions. It could help refine modelling and provide remote assessment of jay locations and of where they spend the majority of their time and potential food resources, and it could also result in less ground work being required to quantify habitat uses (S Somershoe, personal communication 2024). If the pinyon jay is listed as threatened or endangered under the ESA, the results of the random forest model could have an unexpected but beneficial secondary use, informing which areas would probably have the environmental suitability to sustain pinyon pines into the future, which might lead to the planting of new trees. By projecting our model results into the future (SSP5-8.5 for 2081–2100), we saw a 50.31% decrease in the current areas of highest prediction. These results could be beneficial in selecting sites for planting, as environmental factors are likely to be very different in the future from those that exist now, and current locations of highest habitat suitability are not likely to remain the same.

Our modelling area – the northern Great Basin – is vast, and pinyon–juniper woodlands are extremely complex, with nearly every site being different (S Somershoe, personal communication 2024). Anything that can be done to reduce the time, energy and cost of study efforts could have a major impact on the timeliness and effectiveness of such efforts. Refining Landfire's EVT data would be helpful, and by comparing current classifications with our predictions for habitat suitability, areas that could potentially be misclassified as pinyon–juniper woodlands or omitted as such can be highlighted. The pure juniper in the South Hills of southern Idaho, for example (which have no known instances of pinyon pine), has low probability for habitat suitability in our model but has been classified as pinyon–juniper woodland rather than simply juniper. Considering the entire study area, there were 3406.13 km²

of pinyon–juniper woodland classifications that had low probability in our predictions and 16 826.4 km² of high probability for habitat suitability that were not classified as pinyon–juniper woodland. The very strong results for our model indicate its ability to assist in correcting potential misclassifications and adding large amounts of pinyon–juniper woodland that have not yet been classified as such.

Acknowledgements. We would like to acknowledge the Boise State University department of mathematics, specifically Dr Kyungduk Ko, for providing the academic framework to pursue this research question and for the advice. Special thanks go to Wallace Keck (City of Rocks National Reserve superintendent) and Christopher Witt (US Forest Service ecologist) for providing highly accurate pseudo-absence and ground-truthed presence points for pinyon pine.

Financial support. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Competing interests. The authors declare none.

Ethical standards. Due to the nature of the study, ethical permission was not required.

References

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223–1232.

Balda RP, Marzluff JM (1992) *The Pinyon Jay: Behavioral Ecology of a Colonial and Cooperative Corvid*. London, UK: T & AD Poyser.

BBS (2022) BBS trends 1966–2022 (Version Data:core(1966–2022) – Scale: Surveywide – Years: 1966–2022 – Pinyon Jay [*Gymnorhinus cyanocephalus*]) [www document]. URL <https://www.mbr-pwrc.usgs.gov/index22.shtml>

Behl NJ, Benkman CW (2018) Habitat associations and abundance of a range-restricted specialist, the Cassia crossbill (*Loxia sinesciuris*). *The Condor* 120: 666–679.

BirdLife International (2020) *Gymnorhinus cyanocephalus*: The IUCN Red List of Threatened Species 2020: e.T22705608A17959 2026, International Union for Conservation of Nature [www document]. URL <https://www.iucnredlist.org/species/22705608/179592026>

Breiman L (2001) Random forests. *Machine Learning* 45: 5–32.

Breshears DD, Cobb NS, Rich PM, Price KP, Allen CD, Balice RG, et al. (2005) Regional vegetation die-off in response to global-change-type drought. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15144–15148.

Breshears DD, Myers OB, Meyer CW, Barnes FJ, Zou CB, Allen CD, et al. (2009) Tree die-off in response to global change-type drought: mortality insights from a decade of plant water potential measurements. *Frontiers in Ecology and the Environment* 7: 185–189.

Buck RC, Flores-Renteria L (2022) *The Pinyon Pine Syngameon: Modeling the Past to Understand the Present and Predict the Future*. Doctoral thesis. Riverside, CA, USA: University of California, Riverside.

Burns RM, Honkala BH (1990) Volume 2: hardwoods. In: *Silvics of North America*. Washington DC, USA: US Department of Agriculture, Forest Service.

Clifford MJ, Royer PD, Cobb NS, Breshears DD, Ford PL (2013) Precipitation thresholds and drought-induced tree die-off: insights from patterns of *Pinus edulis* mortality along an environmental stress gradient. *New Phytologist* 200: 413–421.

Cohen J (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37–46.

Cole KL, Fisher J, Arundel ST, Cannella J, Swift S (2008a) Geographical and climatic limits of needle types of one- and two-needled pinyon pines. *Journal of Biogeography* 35: 257–269.

Cole KL, Ironside K, Arundel ST, Duffy P, Shaw JD (2008b) Modeling future plant distributions on the Colorado Plateau: an example using *Pinus edulis*. In: C Van Riper, MK Sogge (eds), *The Colorado Plateau III: Integrating*

Research and Resources Management for Effective Conservation (pp. 319–330). Tucson, AZ, USA: University of Arizona Press.

Estrella P (2022) Petition to list the pinyon jay (*Gymnorhinus cyanocephalus*) as endangered or threatened under the Endangered Species Act [www document]. URL https://defenders.org/sites/default/files/inline-files/2022.4.25_FWS_Listing%20petition_Pinyon%20Jay.pdf

Evans JS, Murphy MA, Holden ZA, Cushman SA (2011) Modeling species distribution and change using random forest. In: CA Drew, YF Wiersma, F Huettmann (eds), *Predictive Species and Habitat Modeling in Landscape Ecology* (pp. 139–159). New York, NY, USA: Springer.

Evans RA (1988) Management of pinyon–juniper woodlands (General Technical Report No. Int. 249) [www document]. URL <https://books.google.co.uk/books?hl=en&lr=&id=ofAHREu5UW4C&oi=fnd&pg=PA1&dq=management+of+pinyon+juniper+woodlands+evans&ots=9zWBuu5PeL&sig=bmeErtZkma4B1isRsalG6ATQKk#v=onepage&q&f=false>

Fleming MD, Hoffer RM (1979) *Machine Processing of Landsat MSS Data and DMA Topographic Data*. Laboratory for Applications of Remote Sensing Symposia Technical Report 062879. West Lafayette, IN, USA: Laboratory for Applications of Remote Sensing.

Florkowski CM (2008) Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clinical Biochemist Reviews* 29(Suppl. 1): S83–S87.

Forcella F (1981) Ovulate cone production in pinyon: negative exponential relationship with late summer temperature. *Ecology* 62: 488–491.

Gottfried GJ, Swetnam TW, Allen CD, Betancourt JL, Chung-MacCoubrey AL (1995) Pinyon–juniper woodlands. In: *Ecology, Diversity, and Sustainability of the Middle Rio Grande Basin*, eds. DM Finch, JA Tainter, pp. 95–132. Fort Collins, CO: Rocky Mountain Forest and Range Experiment Station.

Hajian-Talaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine* 4: 627–635.

Han H, Guo X, Yu H (2016) Variable selection using mean decrease accuracy and mean decrease Gini based on random forest. In: *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 219–224). Piscataway, NJ, USA: IEEE.

Hartsell JA, Copeland SM, Munson SM, Butterfield BJ, Bradford JB (2020) Gaps and hotspots in the state of knowledge of pinyon–juniper communities. *Forest Ecology and Management* 455: 117628.

Hernandez PA, Graham CH, Master LL, Albert DL(2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773–785.

Johnson K, Balda RP (2020) Pinyon jay (*Gymnorhinus cyanocephalus*). In: SM Billerman, BK Keeney, PG Rodewald, TS Schulerberg (eds), *Birds of the World*. Ithaca, NY, USA: Cornell Lab of Ornithology.

Johnson K, Sadoti G (2023) A review of pinyon jay (*Gymnorhinus cyanocephalus*) habitat ecology. *Wilson Journal of Ornithology* 135: 232–247.

Kaufmann MR, Graham RT, Boyce DA Jr, Moir WH, Pery L, Reynolds RT, et al. (1994) *An Ecological Basis for Ecosystem Management*. Gen. Tech. Rep. RM-GTR-246. Fort Collins, CO, USA: US Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station.

Kral R (1993) *Pinus linnaeus* [www document]. URL http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=125519

Landfire (2022) About Landfire [www document]. URL <https://landfire.gov/about.php>

Lanner RM (1981) *The Pinyon Pine: A Natural and Cultural History*. Reno, NV, USA: University of Nevada Press.

Mandrekar JN (2010) Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology* 5: 1315–1316.

McDowell NG, Williams AP, Xu C, Pockman WT, Dickman LT, Sevanto S, et al. (2016) Multi-scale predictions of massive conifer mortality due to chronic temperature rise. *Nature Climate Change* 6: 295–300.

McPherson JM, Jetz W, Rogers DJ (2004) The effects of species’ range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? Species’ range and distribution model accuracy. *Journal of Applied Ecology* 41: 811–823.

Meddens AJH, Hicke JA, Macalady AK, Buotte PC, Cowles TR, Allen CD (2015) Patterns and causes of observed piñon pine mortality in the southwestern United States. *New Phytologist* 206: 91–97.

- Meeuwig RO (1979) *Growth Characteristics of Pinyon-Juniper Stands in the Western Great Basin*. USDA Forest Service research paper INT, 238. Washington, DC, USA: USDA.
- Mi C, Huettmann F, Guo Y, Han X, Wen L (2017) Why choose random forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5: e2849.
- Miller EL, Meeuwig RO, Budy JD (1981) *Biomass of Singleleaf Pinyon and Utah Juniper*. US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. Research paper INT-273. Washington, DC, USA: USDA.
- Miller RF, Chambers JC, Evers L, Williams CJ, Snyder KA, Roundy BA, Pierson FB (2019) *The Ecology, History, Ecohydrology, and Management of Pinyon and Juniper Woodlands in the Great Basin and Northern Colorado Plateau of the Western United States*. US Department of Agriculture, Forest Service, Rocky Mountain Research Station. Gen. Tech. Rep. RMRS-GTR-403. Washington, DC, USA: USDA.
- Montes JR, Peláez P, Willyard A, Moreno-Letelier A, Piñero D, Gernandt DS (2019) Phylogenetics of *Pinus* subsection *Cembroides* Engelm. (Pinaceae) inferred from low-copy nuclear gene sequences. *Systematic Botany* 44: 501–518.
- Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91: 252–261.
- NatureServe (2017) *International Ecological Classification Standard: Terrestrial Ecological Classifications*. Arlington, VA, USA: NatureServe Central Databases. Data current as of 28 November 2017.
- Nielsen J (2009a) Pinyon-juniper woodlands – introduction & distribution [www document]. URL <https://www.nps.gov/articles/pinyon-juniper-woodlands-distribution.htm>
- Nielsen J (2009b) Series: pinyon-juniper woodlands [www document]. URL <https://www.nps.gov/articles/series.htm?id=0216D798-933C-2108-EB4384D97499E89A>
- Noel AR, Shriver RK, Crausbay SD, Bradford JB (2023) Where can managers effectively resist climate-driven ecological transformation in pinyon-juniper woodlands of the US Southwest? *Global Change Biology* 29: 4327–4341.
- Parmenter RR, Zlotin RI, Moore DI, Myers OB (2018) Environmental and endogenous drivers of tree mast production and synchrony in piñon-juniper-oak woodlands of New Mexico. *Ecosphere* 9: e02360.
- QGIS Development Team (2023) QGIS Geographic Information System, Open Source Geospatial Foundation Project [www document]. URL <http://qgis.org>
- R Core Team (2023) R: A language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing [www document]. URL <https://www.R-project.org>
- Redmond MD, Forcella F, Barger NN (2012) Declines in pinyon pine cone production associated with regional warming. *Ecosphere* 3: 1–14.
- Romme WH, Allen CD, Bailey JD, Baker WL, Bestelmeyer BT, Brown PM, et al. (2009) Historical and modern disturbance regimes, stand structures, and landscape dynamics in piñon-juniper vegetation of the western United States. *Rangeland Ecology & Management* 62: 203–222.
- Rust SK (1999) Pinyon-juniper woodland classification and description in Research Natural Areas in southeastern Idaho. In: *Ecology and Management of Pinyon-Juniper Communities Within the Interior West* (pp. 82–93). Ogden, UT, USA: US Forest Service.
- Shriver RK, Yackulic CB, Bell DM, Bradford JB, Hampe A (2022) Dry forest decline is driven by both declining recruitment and increasing mortality in response to warm, dry conditions. *Global Ecology and Biogeography* 31: 2259–2269.
- Somershoe SG, Ammon E, Boone JD, Johnson K, Darr M, Witt C, Duvuvuei E (2020) Conservation strategy for the pinyon jay (*Gymnorhinus cyanocephalus*) [www document]. URL <https://partnersinflight.org/resources/conservation-strategy-for-pinyon-jay/>
- Stafford BL, Balda RP, Kamil AC (2006) Does seed-caching experience affect spatial memory performance by pinyon jays? *Ethology* 112: 1202–1208.
- Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1–13.
- Tausch RJ, Hood S (2007) Pinyon/juniper woodlands. In: SM Hood, M Miller (eds), *Fire Ecology and Management of the Major Ecosystems of Southern Utah* (pp. 57–71). Fort Collins, CO, USA: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Van Riper C III, Hatten JR, Giermakowski JT, Mattson D, Homes JA, Johnson MJ, et al. (2014) *Projecting Climate Effects on Birds and Reptiles of the Southwestern United States* (Open-File Report). Washington, DC, USA: US Department of the Interior, US Geological Survey.
- West NE, Tausch RJ, Tueller PT (1998) *A Management-Oriented Classification of Pinyon-Juniper Woodlands of the Great Basin* (No. RMRS-GTR-12). Fort Collins, CO USA: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Williams AP, Allen CD, Millar CI, Swetnam TW, Michaelsen J, Still CJ, Leavitt SW (2010) Forest responses to increasing aridity and warmth in the southwestern United States. *Proceedings of the National Academy of Sciences of the United States of America* 107: 21289–21294.
- Wilson MFJ, O'Connell B, Brown C, Guinan JC, Grehan AJ (2007) Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy* 30: 3–35.
- Wion AP, Weisberg PJ, Pearse IS, Redmond MD (2020) Aridity drives spatiotemporal patterns of masting across the latitudinal range of a dryland conifer. *Ecography* 43: 569–580.
- Yiu T (2019) Understanding random forest – how the algorithm works and why it is so effective [www document]. URL <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>