**Title**: Linking pinyon jay population abundance to pinyon cone production and climate in the SW United States

**Author list**: Ana Miller-ter Kuile1,2, Andreas Wion3, Kyle Rodman4, Jamie S. Sanderlin2

**Author affiliations**:

1 Northern Arizona University School of Informatics, Computing, and Cyber Systems, Flagstaff, Arizona, USA

2 USDA Forest Service Rocky Mountain Research Station, Flagstaff, Arizona, USA

3 USGS Santa Fe, New Mexico, USA

4 Ecological Restoration Institute, Northern Arizona University, Flagstaff, Arizona, USA

**Keywords**:

# Data availability

# Abstract

# 1. Introduction

[big picture]

Co-evolved relationships and how they may be changing with climate change. Additionally, how we may draw understanding about the diversity of responses because these co-evolved relationships occur on the background of other broad-scale variation in species’ ranges that might mediate (good or bad, helping/hurting) the co-evolution and its response to global change drivers.

[zooming in]

In the SW, there is an important co-evolution between pinyon pine and pinyon jays. Jays have been declining over the last several decades. This jay caches pinyon seeds across the landscape to eat or to forget later and help the tree disperse to new areas. Pinyon pines have also seen dramatic shifts in range (expansions, die offs) and reproduction due to climate change and climate variability. [Some biology about pinyon masting. Some more info on the nature of the co-evolution here too]

[gap]

[What is causing their decline?] Given the complex co-evolution of the interaction between the pinyon pine and the pinyon jay, it is hard to tease apart the cause and effect of these patterns or how changes in resource availability alter jay abundances. This has not been evaluated to date. It is likely that, given the broad geographical range of both species, that a different combination of forces may be playing out across the species’ range, and that these effects may not be concurrent. Pinyon pine cone production occurs ~2 years following favorable climate years. Other food resources for jays in these systems may respond more quickly to the same climate conditions (e.g., insects, junipers, and gambel oak). Thus, it may be that jay populations increase \*prior\* to good cone years for pinyon pines as they are either signaled to “get ready” by the same climate conditions as trees or because their populations get a boost prior to a good cone year from other food resources. The result of jays “predicting” good cone years is that more jays will be in the landscape to distribute seeds to more new places. In addition or conversely, jays could also respond with an immediate or lagged effect to cone production. Teasing these patterns out and how other climatic conditions mediate these effects could be imperative for understanding the link between patterns in resources and birds and help in conservation of the species.

[our study]

In this study, we aimed to determine the relationship between pinyon pine cones and pinyon jay abundance in the Southwest US and how other site-level factors may mediate this relationship. To leverage the broadest spatial and temporal availability of data on pinyon jay abundance, we used eBIRD (Cite) data into a model estimating latent abundance (CITE). We then evaluated how pinyon jay abundance related to cone production (CITE andreas), pinyon pine basal area, climate (maximum temperature, precipitation, and monsoonality), and the interactions between cone production and these other variables as broad-scale mediators of the relationship between cone availability and alternative food sources in a given site and a given year (e.g., seeds, acorns, insects). We allowed cone availability to have immediate, lead (jays “predict” cone years), and lag (jays “respond to” cone years) effects on jay abundance using a stochastic antecedent modeling approach (CITE). Relatedly, we allowed maximum temperature and precipitation to have lagged and immediate effects on jay abundance. Specifically, in this study, we wanted to evaluate 1) how strong is the relationship between pinyon jay abundance and cone availability? and 2) how is this relationship mediated by habitat and climate variables that vary by site and/or year? This study provides a valuable way of determining the strength of co-evolution between pinyon pines and pinyon jays and valuable insights into the broad-scale mechanisms by which pines and jays may be declining in the Southwest US and what might be mediating this decline.

# 2. Methods

### 2.1 Study site and pinyon jay biology

Pinyon jays are a widely-spread species that occurs across the western US (California, Nevada, Utah, Arizona, Colorado, New Mexico). The core of their range is in the Great Basin region of Nevada. Pinyon jays are generalist seed- and insect-eaters, but have co-evolved with pinyon pine. They cache [give some info on caching behavior and co-evolution here]. Pinyon jays often occur in flocks. According to several studies [cite here], pinyon jays have been in decline across their range, likely for a variety of reasons [provide some possibilities here]. Concurrently, pinyon pines have seen shifts in their range and reproduction [provide some details here].

Although pinyon jays occur across the western US, in this study, we focused on the intermountain and southwest regions of the US because of available information on cone production (e.g., data from Utah, Colorado, New Mexico, and Arizona; CITE ANDREAS HERE). [Any other justifications for including only data from this region WRT climate/habitat? Please add tree guys].

### 2.2 Bird dataset descriptions

To estimate pinyon jay abundance, we used data from both Breeding Bird Survey (BBS) [cite] and eBIRD [cite] in combination in this study since eBIRD has broader spatial coverage with the environmental covariates we were interested in linking to pinyon jay abundances. For both datasets, we filtered pinyon jay observations to areas with pinyon pine vegetation cover (where did Andreas get the data to mask cone production? CITE).

BBS surveys have occurred since 1966 with the goal of determining long-term trends in bird populations in North America. BBS data are collected each year in June at a set of roadside sampling transects. Each transect is roughly 24.5 miles long with a set of fifty stops roughly 0.5 miles apart. At each stop, observers conduct a 3-minute point count for all birds seen and heard within a 0.25 mile radius. All surveys occur within five hours of local sunrise (CITE). We selected BBS data for pinyon jays from four states (Colorado, Utah, New Mexico, and Arizona) and then masked observations to include only those with starting points within pinyon pine habitat. We also only include data from the first 10 stops (~5 miles; ~8km) in our models because these represent stops within a spatial resolution that is fine enough to mesh with 4x4 km resolution covariate data (described below SECTION ##).

eBIRD is community science data, meaning that non-scientist birders can collect eBIRD data and provide it to the eBIRD database. We followed filtering methods from the eBIRD database curators (Cornell Lab CITE/LINK HERE). Specifically, we selected data that represented only complete checklists (e.g., observers recorded all birds of all species observed, not just individuals of species of interest). We selected checklists from a similar timeframe as BBS data (June and July) and included only observations of “stationary” and “traveling” protocols. We only included checklists that were collected in under five hours, 0-5 km in distance, and with ten or fewer observers.

eBIRD data is unstructured data without a regular sampling design, so we had to spatially sub-sample eBIRD data to limit spatial bias. To do this, we implemented a modified version of a previously-used method that links checklists with and without observations of a species of interest within a certain spatial scale as repeat measurements at that spatial scale (e.g., raster grid cell) (Schindler et al., 2022). We adapted the method in that study; however, to not explicitly select checklists within specific raster grid cells. Rather, we linked observed-not observed checklists based on spatial proximity (1km distance or less). We then spatially subset these pairs along with all checklists (observed and not observed) such that within a given year, all checklists (or checklist pairs) were at least 4km from another checklist (or pair). The coordinates for a pair became their midpoint; coordinates for single observed-not observed checklists remained the same.

### 2.3 Exploring relationships between cone production and jay abundances using BBS from 1960-2023

[Kyle please fill this part in]

### 2.4 Examining recent (2010-2023) relationships between pinyon jay abundance and cone production

To explore relationships between pinyon jay abundance and environmental covariates, especially cone production, we implemented a process of data integration (Miller et al., 2019) using a joint-likelihood approach (Schindler et al., 2022) combining recent (2010-2023) data from both BBS and eBIRD. We chose to do only recent relationships since this is the period in which there are greater amounts of data with which to explore environmental relationships.

For this modeling approach, we created a two-part model with an observation process separately for each of the two bird datasets (BBS and eBIRD) and a biological process model that combines information from both bird dataset observation processes with equal weights to estimate latent “abundance” values. The observation process models include covariates that describe variation in detection probabilities (e.g., observer experience and/or survey conditions). The biological process includes a likelihood that describes the biological drivers of latent abundance. These covariates include a stochastic antecedent modeling structure that allows cone abundance to have both lead and lagged effects (e.g., birds can respond before, concurrent, or after high cone years) and other climate variables to have lagged effects (Ogle et al., 2015). We estimated latent abundance at a 4x4 km scale because this is the scale of the spatial covariates. We describe how we account for spatial uncertainty in the bird observation datasets (i.e. “downscale”) in section \_\_\_ below.

### 2.4.1 Covariate datasets

[explanations of covariate datasets]

### 2.4.2 Observation process for BBS and eBIRD

In a joint-likelihood model, multiple independent datasets contribute to a latent “true” value for a variable. In this model, we are combining observed counts from BBS and eBIRD to estimate a “true” latent abundance at each grid, , in each year , , of pinyon jays in a given 4x4 km grid cell in a given year. Both of these observation datasets are part of separate observation process models that incorporate covariates that impact the detection of birds within each survey. For both observation processes, observed count data are binomially distributed based on a detection probability for a site in a year and the “true” latent abundance, , being the total possible “trials”.

BBS surveys are conducted on a set of transects. In this study, we have subset the first 10 stops on any transect - these stops are considered spatial replicates for that transect and all contribute to one estimate of detection probability for that transect. We considered observer experience (in years) as a covariate to detection of pinyon jays for BBS surveys. The transects are located in a 4x4 km grid cell and transect-level information is linked back to this grid cell based on (more information about the process for this below \_\_\_). BBS count data for each transect, , year, , and stop, , is distributed as:

The detection probability for BBS data includes a regression with the detection covariate of observer experience (in years), which depends on site and year, but not on stop (the same observer conducted all stops on a transect in a year):

eBIRD surveys are conducted at a set of points with a set survey distance. We spatially sub-sampled eBIRD observations (described above) so that pairs of observed-absent checklists or single checklists of either type were 4 km from each other. We buffered checklists so that none occurred within 10km of the “edge” of the four states for which we had cone abundance data (Utah, Colorado, New Mexico, and Arizona). Again, eBIRD observed count data for pair, , in year, , in checklist, , are binomially distributed based on a detection probability and the latent true abundance of birds in the 4x4 km grid cell in which the checklist is located:

The detection probability for eBIRD data includes a regression with the detection covariates of survey type (stationary or traveling), start time, duration of survey, distance traveled, and the number of observers for each checklist. See above (\_\_\_) for how we filtered these variables initially based on best practices so that checklists represented similar effort. These covariates all depended on group, year, and checklist:

For both datasets, we wanted to account for spatial uncertainty in both the coordinates provided in the datasets as well as where within each survey pinyon jays were actually detected (e.g., we did not know if jays were detected at the beginning of the survey or halfway into the survey, or the end of a survey - a distance that could be up to 8km away from the starting location). To account for this spatial uncertainty, we determined which 4x4 km grid cell each survey (BBS transect or eBIRD checklist pair) was located in based on a stochastic process that depended on what proportion of a buffered radius around each survey was contained within a given grid cell. We sampled from possible grid cells using these proportions as a vector of probabilities for each transect or pair and year, , of birds being located in that grid cell based on a categorical distribution. Thus, for each iteration, , of the model, the grid cell ID for either eBIRD or BBS is given as:

### 2.4.3 Biological process of latent pinyon jay abundance with covariates with lead and lagged effects

As highlighted in the section above, the joint observation processes contribute to a shared estimate of “true” latent abundance, , of pinyon jays in a given 4x4 km grid cell in a given year. is Poisson distributed around an expected rate parameter, , as follows:

Furthermore, is dependent on a regression of environmental covariates:

In this formulation, covariates include only concurrent effects. These variables in the model include a cell-level monsoonality score and pinyon pine basal area (yearly). Conversely, values for each comprise a stochastic antecedent structure (Ogle et al., 2015) dependent on the covariate, where each is the weighted average value of a set of seasonal or yearly values for that covariate across a defined set of seasons or years. The weights for each covariate across all seasons or years, *l*, , have a Dirichlet prior and sum to 1. The weight, or importance weight, for each season or year indicates how important that value is to the overall effect (larger values of the weight mean more importance of that season or year to that covariate effect). The covariates for which antecedent weights are calculated are cone production, maximum temperature, and precipitation. Cone production is considered on a yearly scale (cones only produce up to one time in a year) and include both lead and lagged effects (e.g., jays “predict” cone years or “respond” to good cone years), and include \_\_\_ years prior years, the current year, and \_\_ years after the current year. Both temperature and precipitation are at a seasonal scale and only consider lagged effects. The seasons are based on important seasonal biology for pinyon jays, including: breeding (February-April), feeding dependent young (May-June), summer (July), fall and winter foraging, potentially with irruption (August - January) (Wiggins, D. A., 2005). For seasonal covariates, we considered \_\_ seasons into the past. We also considered interactions between cone production and all other covariates (pinyon basal area, monsoonality, maximum temperature, and precipitation) to determine how the relationship between cones and birds is mediated by other covariates related to cone reliability and alternative food resources (e.g., juniper berries, acorns, insects).

| Covariate or interaction | Description (all are at 4x4 km grid cell scale) | Question |
| --- | --- | --- |
| Pinyon pine cone abundance | Abundance of cones from a predictive model (CITE) | How is pinyon jay abundance linked to pinyon cone availability (can have lead, immediate, and lagged effects)? |
| Maximum temperature | A proxy for physiological limits and/or insect resource availability (can have lagged effects) | How is pinyon jay abundance linked to physiological limits and/or insect availability? |
| Cumulative precipitation | A proxy for insect resource availability (can have lagged effects) | How is pinyon jay abundance linked to insect availability? |
| Monsoonality | The degree to which the site depends on monsoon moisture | How is pinyon jay abundance shaped by broad-scale climatic mediators of cone reliability? |
| Pinyon basal area | The basal area of pinyon pine | How is pinyon jay abundance shaped by local habitat preferences for food trees? |
| Cone abundance x Temperature |  | Is the importance of cone abundance for jays mediated by temperature? |
| Cone abundance x Precipiation |  | Is the importance of cone abundance for jays mediated by precipitation (e.g., more/less availability of insect prey) |
| Cone abundance x Monsoonality |  | Is the importance of cone abundance for jays different across a range of monsoonal conditions that alter cone reliability? |
| Cone abundance x Pinyon basal area |  | Is the importance of cone abundance for jays different depending on the local habitat configuration and abundance of pinyon pine? |

### 2.4.4 Model implementation, convergence, and diagnostics

We implemented models in the Bayesian software JAGS (CITE) using R (CITE) and the jagsUI wrapper package (CITE). We prepared data using the here (CITE), tidyverse (CITE), sf (CITE), terra (CITE), readxl (CITE), sf (CITE), exactextractr (CITE), spatialEco (CITE), nngeo (CITE), sp (CITE), auk (CITE), lubridate (CITE), prism (CITE), and data.table (CITE) packages.

We initially ran models to generate 4000 MCMC samples with which to determine how many iterations would be needed to reach convergence (CITE). We then re-ran models with more iterations and initial values sufficient to reach convergence based on the raftery.diag() function in the coda package (CITE). We assessed model convergence by evaluating trace, history, and autocorrelation plots generated with the mcmcplots package (). We quantified model convergence by evaluating whether Gelman-Ruman statistics, , for all root nodes in the model were < 1.2. We assessed model goodness-of-fit by comparing observed count data from eBIRD and BBS to predicted count data for each dataset based on the model mean value.

# Results

# Discussion

# Tables and Figures

# References

# 

Miller, D. A. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species’ distributions. *Methods in Ecology and Evolution*, *10*(1), 22–37. <https://doi.org/10.1111/2041-210X.13110>

Ogle, K., Barber, J. J., Barron-Gafford, G. A., Bentley, L. P., Young, J. M., Huxman, T. E., Loik, M. E., & Tissue, D. T. (2015). Quantifying ecological memory in plant and ecosystem processes. *Ecology Letters*, *18*(3), 221–235. <https://doi.org/10.1111/ele.12399>

Schindler, A. R., Cunningham, S. A., Schafer, T. L. J., Sinnott, E. A., Clements, S. J., DiDonato, F. M., Mosloff, A. R., Walters, C. M., Shipley, A. A., Weegman, M. D., & Zhao, Q. (2022). Joint analysis of structured and semi-structured community science data improves precision of relative abundance but not trends in birds. *Scientific Reports*, *12*(1), 20289. <https://doi.org/10.1038/s41598-022-23603-0>

Wiggins, D. A. (2005). *Pinyon jay (gymnorhinus cyanocephalus): A technical conservation assessment.* [http://www.fs.fed.us/r2/projects/scp/ assessments/pinyonjay.pdf](http://www.fs.fed.us/r2/projects/scp/%20assessments/pinyonjay.pdf)