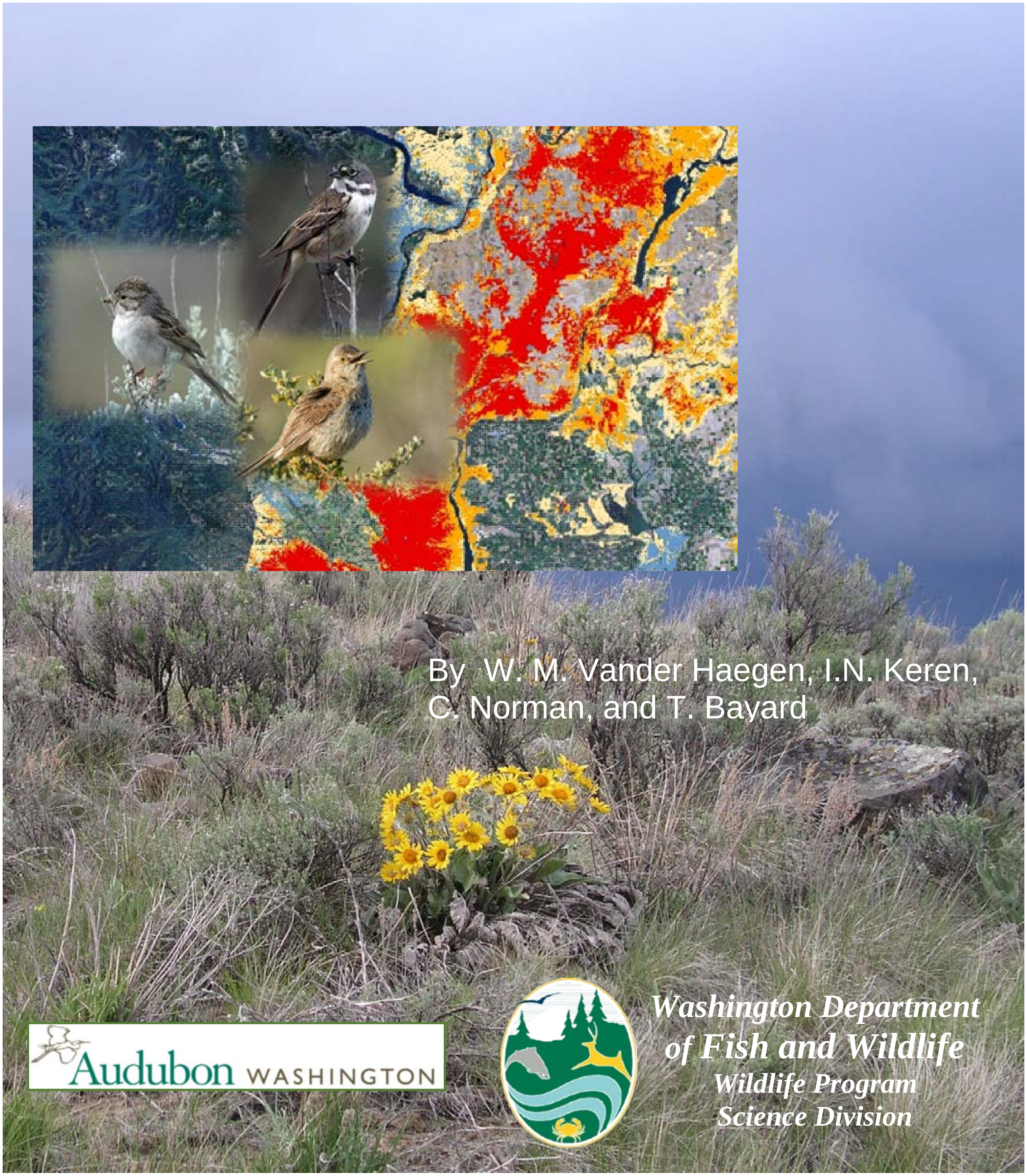
Occurrence and predicted distribution of sagebrush obligate passerines in Washington State: a community-science based project



By W. M. Vander Haegen, I.N. Keren, C. Norman, and T. Bayard

Audubon WASHINGTON

Washington Department of Fish and Wildlife
Wildlife Program
Science Division

Recommended Citation:

Vander Haegen, W. M., I. N. Keren, C. Norman, and T. Bayard. 2022. Occurrence and predicted distribution of sagebrush obligate passerines in Washington State: a community-science based project. Final Report. Washington Department of Fish and Wildlife, Olympia.

Cover photos: Bird photos by Ivar Husa (sagebrush sparrow), Mick Thompson (sage thrasher), and Matt Vander Haegen (Brewer's sparrow). Landscape photo by Matt Vander Haegen.

Occurrence and predicted distribution of sagebrush obligate passerines in Washington State: a community-science based project

W. Matthew Vander Haegen and Ilai N. Keren

Washington Department of Fish and Wildlife
Wildlife Program, Science Division

Christi Norman and Trina Bayard

Audubon Washington

08 June 2022

Final Report

Washington Department of Fish and Wildlife

Wildlife Program, Science Division

1111 Washington Street

Olympia, WA 98501

ACKNOWLEGEMENTS

This project was a collaborative effort between the Washington Department of Fish and Wildlife (WDFW), Audubon chapters, Audubon Washington (AW), and the National Audubon Society. It would not have been possible without the dedication and tremendous effort put forth by members of numerous Audubon chapters in Washington, both in the field and in numerous, less exciting office tasks. The Lower Columbia Basin Chapter was instrumental in helping fine-tune the approach and methods during the pilot year of 2014. We deeply appreciate the dedication of all Audubon Chapter members and other volunteers who participated in the project. Among those deserving special thanks are Kathy Criddle, Jan Demorest, Cindy Easterson, Catherine Flick, Lindell Haggin, Mark Johnston, Steve Moore, Robbin Priddy, Richard Scranton, Amanda Sherling, Marilyn Sherling, and Lori Wollerman-Nelson. Technology in its various wonderful and frustrating forms was integral to this project. We thank Brian Cosentino and Bill Simper (WDFW), Ryan Hobbs (AW), and the staff at ESRI Olympia for assistance with geographic information systems, global positioning systems, and for designing data entry platforms. Identifying and contacting willing landowners was a critical part of our survey efforts; we thank the many WDFW Private Lands Biologists and District Wildlife and Habitat Biologists for their outstanding efforts. We thank Mike Schroeder for providing a very helpful review of a previous draft of this report. And finally, we dedicate this report to John Pierce, former Chief Scientist with WDFW, for his unwavering support of community science as a means to further conservation of our natural resources.

EXECUTIVE SUMMARY

The shrubsteppe landscapes of eastern Washington have changed significantly in the last century and continue to change in ways detrimental to sagebrush obligate birds such as sagebrush sparrows (*Artemisiospiza nevadensis*), sage thrashers (*Oreoscoptes montanus*), and Brewer's sparrows (*Spizella breweri*). Primary data gaps for conservation of these species include a better understanding of where they occur in eastern Washington and what features they are associated with on the landscape. From 2014-2019 we conducted a community science project to expand our knowledge of where passerine bird species associated with shrubsteppe systems currently occur in the state, with the goal of producing information useful to their conservation. We used a combination of point counts and walking surveys to survey birds at 324 sites and developed empirical models of species distribution based on observed habitat relationships. We surveyed each site 3 times in a single year between April and June with visits spaced >2 weeks apart. Observers entered all observations from surveys conducted on public land into the online database eBird and entered observations from surveys conducted on private lands into a stand-alone database not available to the public to protect landowner privacy. We used 4 different statistical models to examine relationships between species presence and site variables and combined these using an ensemble modeling approach to create species distribution models. A final prediction layer was then calculated as the unweighted mean of the 4 models. These final prediction layers mapped the probability of the species occurring in each pixel (180m) across the landscape, with values ranging from 0.05 to 0.95. We developed model concordance layers that attributed each pixel with the number of models that predicted presence, using individual cut points derived for each model. In the concordance layer, each pixel had a value ranging from 0 to 4, with higher values representing increasing evidence that the species might be present as estimated by the models. We provide guidance on the use of these model concordance layers for conservation planning. Audubon chapters from across the state participated in the surveys, with 7 eastern Washington chapters taking the lead in establishing survey sites, training surveyors, coordinating surveys, and entering and proofing data. Over the 6 years of the project over 285 volunteers (mostly Audubon members) participated in the surveys, completing 987 individual surveys and logging over 14,000 volunteer hours.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

The shrubsteppe landscapes of eastern Washington have changed significantly in the last century and continue to change in ways detrimental to sagebrush obligate birds (Vander Haegen et al. 2000, Knick et al. 2003). Sagebrush sparrows (*Artemisiospiza nevadensis*), sage thrashers (*Oreoscoptes montanus*), and Brewer's sparrows (*Spizella breweri*) all are considered sagebrush obligates (Braun et al. 1976) and the sagebrush sparrow and sage thrasher are listed as species of greatest conservation need in Washington (WDFW 2015). While none of these species currently are listed as threatened or endangered (federally or by the State of Washington), all depend on shrubsteppe habitats that continue to be endangered by conversion to agriculture and suburban/urban development, energy projects such as wind and solar, fragmentation, and vegetation change brought about by wildfire, invasive species, and a changing climate (Knick et al. 2003, Wisdom et al. 2005). A primary data gap for conservation of these species is a better understanding of where they occur in eastern Washington and what features they are associated with on the landscape.

Existing information on bird distribution in Washington can be found in the Washington breeding bird atlas, an effort that began in 1987 and included observations collected up to the mid-1990s (Smith et al. 1997). While the atlas represents an extraordinary effort for the time, the observations used in its development are now dated and most were not collected using standardized survey methods. More recent observations can be found in the online data base eBird (ebird.com), an effort that captures bird observations entered by birders across the world and that includes many observations for the species considered here. While the observations in eBird are proving to be of great value to science (Sullivan et al. 2014), they did not meet our requirements for a rigorous modeling of species distribution; i. e., observations from ad hoc surveys tend to be biased to areas that are easily accessed or popular, species that are most conspicuous, and often span multiple decades (Rondinini et al. 2006). Understanding the current distribution of a species requires unbiased sampling of the area of interest using a random or systematic approach conducted within the time period of interest (Pollock et al. 2002). Information on species occurrences obtained by surveys conducted at specific sites (e.g., point-count observations) has been used to examine important aspects of a species' ecology including range limits, migration timing (here, eBird data have been very useful), and resource selection, but are limited in what they can tell us about distribution on the landscape. Species distribution models (SDMs) make it possible to expand the inference of wildlife point observations to an area much larger than the individual sites surveyed (Guisan and Zimmermann 2000). This is immensely important to conservation planning, as it decreases the omission error (attributing a site as not suitable for a species when it actually is suitable) when applying data to large areas and vastly increases options available to planners (Rondinini et al. 2006).

We began a community science project in 2014 to expand our knowledge of where passerine bird species associated with shrubsteppe systems currently occur in the state, with the goal of producing information useful to their conservation. While there have been many scientific studies of habitat associations and effects of anthropomorphic changes in Washington's shrubsteppe on sagebrush-obligate birds (e.g, Dobler et al. 1994, Vander Haegen et al. 2000, 2015, Earnst and Holmes 2012, Millikin et al. 2020), we lacked a broad view of where these species occur across the state. Specific objectives of the project were to 1) conduct presence/absence surveys for species of conservation concern at a systematic set of locations across Washington's shrubsteppe ecosystems, 2) develop empirical models of species distribution based on observed habitat relationships, and 3) engage community science in planning, data collection, and overall project management, empowering local Audubon Chapters to aid directly in species conservation. In addition to the 3 focal shrubsteppe obligate species mentioned above, observations of 3 passerines associated with shrubsteppe systems (grasshopper sparrow [*Ammodramus savannarum*], savannah sparrow [*Passerculus sandwichensis*], and vesper sparrow [*Pooecetes gramineus*]) also were recorded but are not the focus of this report.

STUDY AREA

The project was centered in the Columbia Plateau Ecoregion and encompassed most of the semi-arid steppe and sagebrush-steppe of eastern Washington (hereafter, shrubsteppe). At least half of the historical shrubsteppe within the study area had been converted to agriculture and other uses (Jacobson and Snyder 2000). Federal and state agencies managed large areas of unconverted lands including the Yakima Training Area (Department of Defense), the Hanford Nuclear Reservation (Department of Energy and Department of Interior), and many state wildlife areas and other state lands managed by the Department of Fish and Wildlife and the Department of Natural resources. Land ownership within the area used to survey for birds and to model species distribution was 68.9% private, 11.8% federal government, 10.8% tribal government, 8.1% state government, and <1% County, municipal, and "other".

METHODS

Sampling Frame

We used data layers of ecological systems defined by NatureServe (http://www.natureserve.org) as a spatial representation of unconverted shrubsteppe in Washington. We compiled ecological systems strata from LANDFIRE 2016 Remap (https://www.landfire.gov/index.php, accessed on 10/20/2018) at a spatial resolution of 30m. LANDFIRE 2016 Remap used Landsat imagery from 2013-2017 and prioritized imagery from 2016. We selected 12 ecological systems from the existing vegetation type (EVT) layer (Table 1) that comprise vegetation communities dominated by native bunchgrasses and, where woody vegetation is included, this component was largely shrubs (e.g. Artemisia spp.). Where EVT was coded as a disturbed type (e.g., Recently Burned Herb and Grass Cover), we substituted the vegetation class from LANDFIRE 2016 Biophysical Setting layer for those pixels when that layer indicated one of the 12 ecological systems selected above.

We created a 1km^2 resolution grid that spanned the shrubsteppe layer to facilitate systematic selection of survey sites. To space potential survey sites across the landscape, we grouped grid cells into 5x5 squares and selected the top-right cell in each square. This provided 2425 potential survey cells, each 1km^2 in area. We intersected these cells with the shrubsteppe layer and calculated the proportion of each cell represented by shrubsteppe. A total of 1633 cells with a minimum 5% shrubsteppe were carried forward for further analysis (Fig. 1). We intersected these cells with a public lands layer to

Table 1. Ecological systems (from Landfire 2016) used to spatially represent shrubsteppe communities in eastern Washington for modeling shrubsteppe bird distributions.

Columbia Basin Foothill and Canyon Dry Grassland
Columbia Basin Palouse Prairie
Columbia Plateau Low Sagebrush Steppe
Columbia Plateau Scabland Shrubland
Columbia Plateau Steppe and Grassland
Inter-Mountain Basins Big Sagebrush Steppe
Inter-Mountain Basins Greasewood Flat
Inter-Mountain Basins Montane Sagebrush Steppe
Inter-Mountain Basins Semi-Desert Shrub Steppe
Great Basin & Intermountain Introduced Annual and Biennial Forbland
Great Basin & Intermountain Introduced Annual Grassland
Great Basin & Intermountain Introduced Perennial Grassland and Forbland

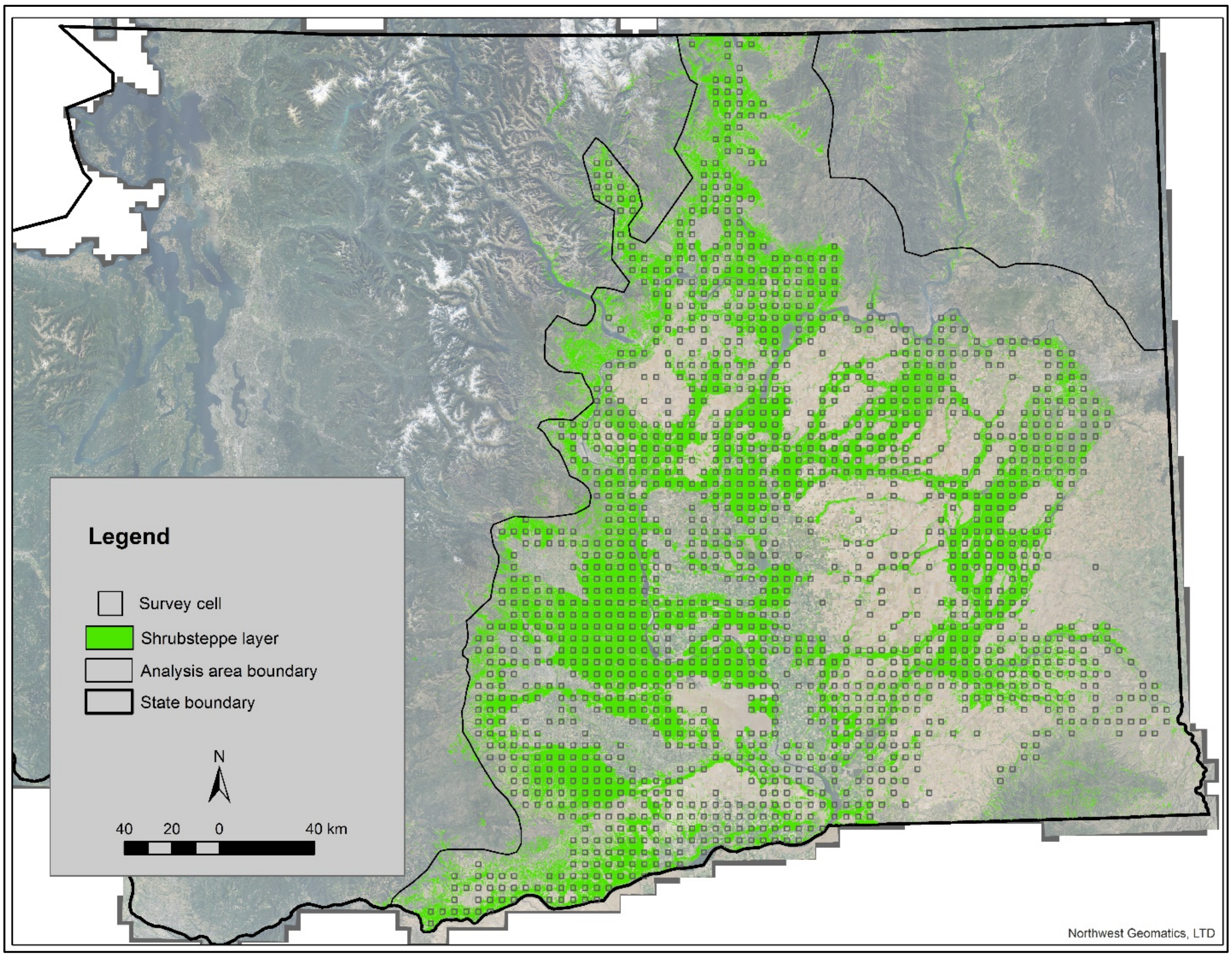


Figure 1. Spatial extent of mapped shrubsteppe in eastern Washington and potential 1km^2 survey cells for study of shrubsteppe bird distribution, 2014-2019.

identify polygons of shrubsteppe on public lands that might be suitable candidates for survey. Because large expanses of public land (e.g., Yakima Training Center, Hanford Nuclear Reservation) contained many qualifying cells, we systematically selected every 3rd cell from the original selection layer within large ownerships to avoid oversampling these extensive areas. To include private lands in our survey, we also intersected the cells with layers representing lands enrolled in WDFW Private Lands Hunting Opportunity program and in the USDA SAFE-Conservation Reserve Program (Fig. 2). Lands enrolled in these programs were known to WDFW regional biologists who facilitated gaining landowner permission. Intersecting the cells that met the 5% minimum with the various land ownership categories yielded 424 potential survey sites on public lands and 108 on private lands. We did not survey sites on Tribal Government lands (Fig. 2).

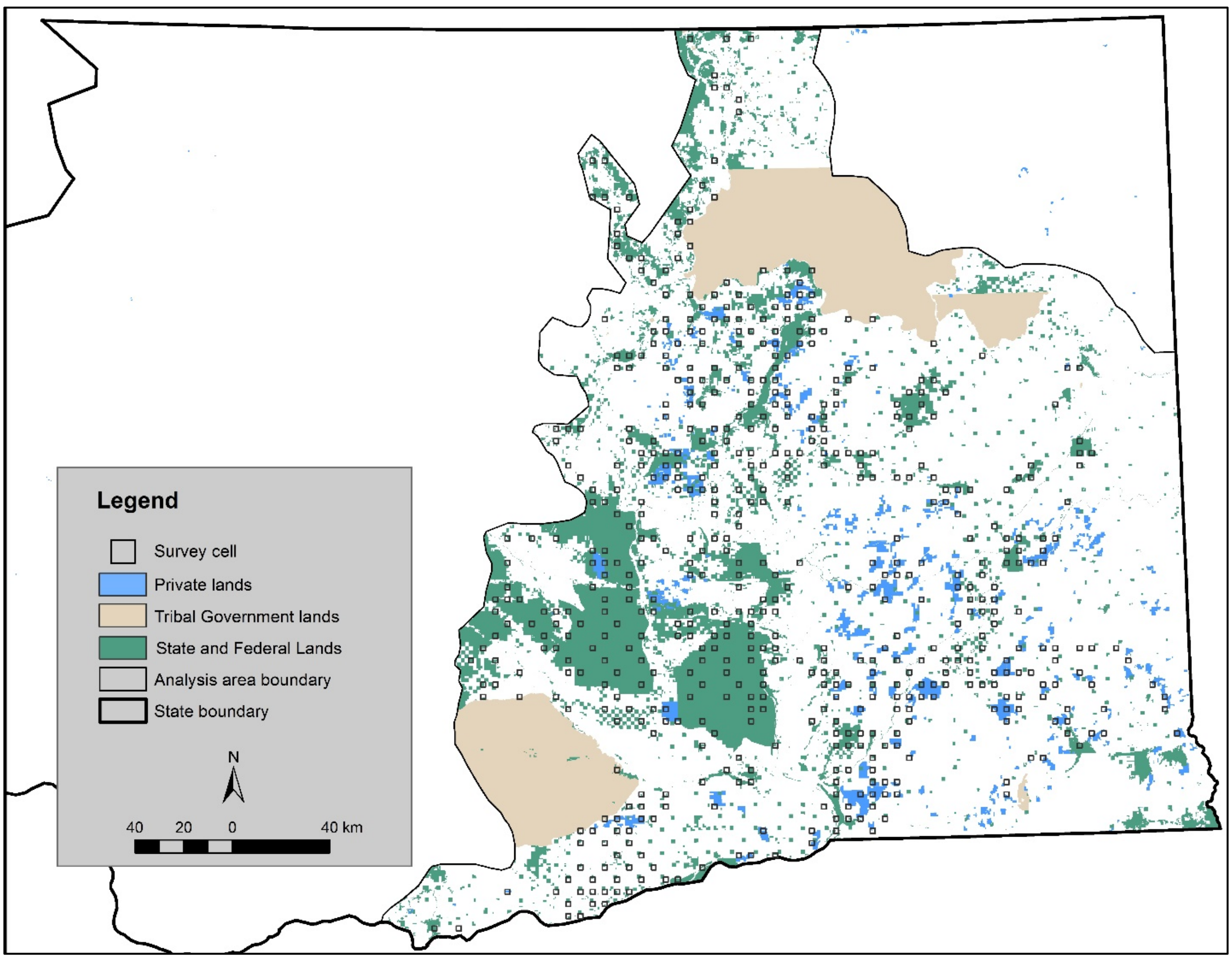


Figure 2. Land ownership categories and final set of potential 1km^2 survey cells for study of shrubsteppe bird distribution in Washington, 2014-2019. Private lands were limited to those enrolled in state or federal programs that facilitated entry by surveyors (see text for details).

After screening the available sites (see Site Screening), we expanded the cell size to 4km^2 in order to increase the area available for survey across all landowner types. We chose to expand the size of cells in the existing pool rather than add additional 1km^2 cells to retain the original, systematic approach and cell spacing. We designated the largest qualifying shrubsteppe polygon in each cell as a potential sampling site and used GIS to place a potential survey point in its geometric center. We used ArcMap Ver. 10.6.1 and ArcPro ver. 2.4.1 (https://www.esri.com) for all spatial data manipulations and analyses.

Site Screening

Potential survey sites in all 532 cells with shrubsteppe on public lands or select private lands were inspected for actual landcover, topography, and potential vehicle access using GIS. We used digital orthophoto layers to inspect vegetation and digital USGS topographic maps and roads layers to assess topography and access. Public ownership layers and county tax layers were used to determine property boundaries. Potential parking locations were identified and marked with the goal of keeping walking distance to the survey point to <1km. All vetting of sites was accomplished by Audubon Chapter members using ArcGIS Online (https://www.esri.com) and data layers mentioned above. During screening, chapter members attributed a digital site layer with information including site name, year, and notes on access and permission.

Potential survey sites identified through GIS screening were visited in the field to confirm suitability of habitat and safety of access. A team of 2-4 individuals from the local Audubon Chapter scouted each site in the fall or spring prior to the survey period and navigated to the potential survey point using a global positioning system (GPS). If the vegetation at the point was not native grassland and shrubs, or if the point fell in topography that was unsafe or not conducive to a point-count survey (e.g., steep slopes or restricted view) the team had the latitude to move the point within the selected shrubsteppe polygon. This was important, because surveying appropriate habitat in each polygon aligned with our value of avoiding false negatives for focal species. Once the final survey point was established, the team recorded the location using GPS and determined the best walking route to a suitable parking location. The team marked the parking location with GPS, and these 2 points (survey and parking) defined the walking survey that was part of the survey protocol (described below). We contacted landowners or land managers of all sites for permission before entering their lands.

Survey Methods

We used a combination of unlimited-radius point counts and walking surveys to survey birds at each site. The point counts provided a standardized, minimum effort repeated among all sites, whereas the walking surveys increased the opportunity to detect focal species and made use of the time surveyors spent walking to and from the point. Walking routes traversed shrubsteppe habitat for most of their length and these added observations reduced the possibility of false negatives for the site. During each survey, observers walked from the parking spot to the survey point, noting all individuals of focal species seen or heard regardless of distance from the walking path. The goal of each survey was to count all observations of the focal species, but observers were free to include any species detected and identified to species. Once arriving at the survey point the observers noted the time and began a 10-min point-count survey. During this survey, the observers noted species detected (as above) regardless of distance from the point. At the conclusion of the point-count, the surveyors retraced their walking route back to the parking spot, again noting all birds seen or heard. During all surveys individual birds were counted along with sex (if determined), and behavior (singing, flying, perched). Results from the 2 walking surveys and the point-count were recorded separately.

We surveyed each site 3 times in a single year with visits spaced >2 weeks apart. We surveyed sites between 1 April and 30 June and between dawn and 0900 hrs. At least one experienced observer and up to 3 additional participants comprised a survey team. Experienced observers were trained in the identification of the focal species by sight and by sound. Sites were surveyed by ≥2 different teams during the season. Sites on public lands where access was restricted (i.e., Yakima Firing Center and Hanford Nuclear Reservation) were conducted by agency biologists using the same protocols, although sites generally were surveyed by the same observer each visit. Observers entered all observations from surveys conducted on public land into the online database eBird (eBird.com). Each site entered into eBird was attributed with a unique location name and a common prefix that allowed us to extract the observations from the eBird database at the end of each season. Observers used ArcGIS Collector to entered observations from surveys conducted on private lands into a stand-alone database not available to the public to protect landowner privacy.

GIS Data Layers for Modeling

We reviewed the literature on shrubsteppe bird habitat and landscape associations, particular work conducting in Washington, and selected 10 explanatory variables to include in our models of species distribution (Table 2). These 10 variables were reported as being associated with occurrence or abundance of the focal bird species in empirical research, or represented metrics closely related to the measured variables. We also included latitude, longitude, and elevation in our models to examine potential spatial patterns in species distribution that were not accounted for in the set of 10 selected variables. Our primary goal in this modeling effort was to predict species distribution, so we were not concerned with including non-biological variables that might prove difficult to interpret.

Landscape Variables

We developed digital layers to spatially represent current shrubsteppe and land enrolled in the Conservation Reserve Program (CRP) on the landscape for use in modeling. These layers made use of the most current information available at a state-wide scale. Shrubsteppe was represented by the 12 ecological systems layers in LANDFIRE described above. Lands enrolled the CRP frequently have vegetation structure similar to that in shrubsteppe and are of value to shrubsteppe birds (Schroeder and Vander Haegen 2011, Vander Haegen et al. 2015). The CRP program in Washington is unique in that sagebrush often is included in the mix of seeds planted on these agricultural lands along with a mix of native and non-native grasses and forbs (Vander Haegen et al. 2015). Sagebrush also frequently expands into CRP from adjacent native shrubsteppe. Because ecological systems in LANDFIRE were mapped using satellite remote sensing, and CRP fields have vegetation (and therefor reflectance values) similar to native shrubsteppe vegetation (Jacobson and Snyder 2000), we masked known CRP fields from the shrubsteppe layer. It was important to identify shrubsteppe and CRP separately for our modeling so that we could examine the influence of each vegetation type independently.

Table 2. Variables used in models of sagebrush bird species distribution in Washington and citations for supporting literature.

Variable Code	Description
Latitude[a]	Latitude or northing within the study area
Longitude	Longitude or easting within the study area
Agsoil	Binary: soil type is agricultural (soil depth coded = 154) or not
Soil depth[a]	Soil depth in inches, where the Agsoil variable is not “True”
Elevation	Land elevation in meters
CRP_3km	Percent landscape in Conservation Reserve Program (CRP) lands within 3km radius
SS_3km[a,b]	Percent landscape in shrubsteppe within 3km radius
SSCRP_3km	Percent landscape in shrubsteppe or CRP within 3km radius
Bare ground[c,d]	Percent cover of bare ground
Herbaceous[a,c,e,f]	Percent cover of herbaceous vegetation
Litter[c]	Percent cover of plant litter
Shrub[a,b,d,f,g]	Percent cover of all shrubs
Sagebrush[b,d,e,h]	Percent cover of sagebrush species

[a] Vander Haegen et al. 2000, [b] Knick and Rotenberry 1995, [c] Rotenberry and Wiens 1980, [d] Larson and Bock 1984, [e] Earnst and Holmes 2021, [f] Vander Haegen et al. 2015, [g] Winter et al. 2006, [h] Wiens and Rotenberry 1981.

To spatially represent CRP fields on the landscape we used a cropland database maintained by the Washington Department of Agriculture (WSDA; https://agr.wa.gov/departments/land-and-water/natural-resources/agricultural-land-use. Accessed 10/6/2020). We used the 2018 WSDA data layer and assigned all crop fields where crop type was designated as “other” to CRP. A visual inspection of a 2006 CRP layer obtained from USDA-Farm Services Agency and the 2018 cropland layer revealed a high degree of spatial overlap in most counties indicating that most current CRP fields would be classified as “other” in the cropland layer. Errors of omission and commission likely were due to fields leaving the CRP program and returning to cultivation after 2006, tilled cropland entering the CRP program after 2006, and active croplands attributed as “other” in the WSDA database. Acreage statistics from the 2018 cropland layer and from an annual report on CRP enrollment statistics from USDA for 2018 (https://www.fsa.usda.gov/Assets/USDA-FSA-Public/usdafiles/Conservation/PDF/December%202018%20Summary.pdf, accessed 8/15/2020) provided additional evidence that the WSDA layer represents fields largely in CRP. Comparing statistics from these 2 sources for the 17 counties encompassed in our modeling effort, the total acreage reported in CRP by USDA and in “other” croplands by WSDA differed by only 2.8%.

We used the shrubsteppe layer and the derived CRP layer to create 9 separate spatial GIS layers, each representing the proportion of these landcover types in the landscape surrounding each survey site and calculated to distances of 1, 3, and 5km from the point. To achieve this, we used a moving window analysis to attribute each pixel with the proportion of the surrounding landscape composed of either shrubsteppe, CRP, or a combination of shrubsteppe and CRP (treating these 2 cover types as one). The combined shrubsteppe-CRP layer was used to examine if birds responded to the proportion of “untilled” land in the landscape surrounding a survey point, treating areas in native shrubsteppe and the structurally-similar CRP as the same. We wondered whether perennial grasslands and shrublands, regardless of their origin (native or planted), created a landscape acceptable to area-sensitive species. Layers with 3km radius captured the most variability among the 3 metrics and were used in all models.

Soil depth and spatial variables

Soil depth was compiled using data from the Natural Resources Conservation Service (NRCS) Soil Survey Geographic Database (SSURGO) and, for areas where SSURGO data were not available, from the U.S. General Soil Map (https://websoilsurvey.nrcs.usda.gov). From these data layers we also created a separate, binary variable to indicate sites classified as cropland in the soil database (coded 152); the primary soil depth variable in these 2 databases contained values only for non-cropland sites. We included a layer for elevation from the digital height mosaic layer developed by the National Land Cover Database (NLCD; https://www.usgs.gov/centers/eros/science/national-land-cover-database) and layers for latitude and longitude as additional spatial predictors.

Fractional Vegetation Layers

We used vegetation data layers in the NLCD to spatially represent vegetation and ground cover characteristics across our study area for use in SDMs. Vegetation layers in the NLCD were developed using a multi-step process beginning with high-resolution (2m) imagery of selected sample sites that were characterized using ground-based vegetation sampling (imagery and ground sampling in 2016) (Rigge et al. 2020). These data were then scaled up using multiple Landsat satellite images (2016 Landsat 8, 30m resolution) and ancillary data (e.g., slope and aspect) to create landscape-scale products (Rigge et al. 2020). We selected 5 layers for vegetation and site parameters know to be associated with the bird species we were modeling and that had suitable cross-validation values (Table 3).

Because the raster data for all GIS layers were specific for a 30m pixel and our bird surveys were designed to document species presence over a larger area, we averaged the individual values for pixels around the survey point. The value of a covariate attributed to a site was the mean of all pixels (n ranged from 8 to 12) intersected by a circle with radius of 60m centered on the site coordinates.

Table 3. **Validation metrics (coefficient of determination [R^2], Root Mean Square Error [RMSE], and slope) for cross- and independent validation of fractional cover layers used in models of species distribution (from Rigge et al. 2020). Variable names use in models are defined in Table 2.**

Layer	Validation (cross, independent)		
	R^2	RMSE	Slope
Shrub	0.73, 0.37	6.0, 10.6	0.70, 0.50
Sagebrush	0.63, 0.4	3.4, 7.5	0.63, 0.52
Herbaceous	0.79, 0.67	6.3, 13.1	0.74, 0.61
Litter	0.75, 0.35	3.8, 8.9	0.71, 0.42
Bare ground	0.85, 0.7	8.0, 14.6	0.78, 0.73

Data Analysis

We compiled observations for the 3 focal species across all surveys to derive presence/absence at each site. Walking surveys averaged 0.52km (range 0.071 to 2.26km) and traversed primarily shrubsteppe vegetation. Combining observations from the walking surveys and point counts increased the number of detections for most species and met our goal of assessing presence/absence for the site as a whole. We addressed detectability of species at survey sites by employing rigorous field methods (3 visits with ≥2 different, experienced observers using 2 survey methods) which maximized probability of detection.

We used 4 different statistical models to examine relationships between species presence and site variables and combined these using an ensemble modeling approach to SDMs (Araujo and New 2007) (Fig. 3). The ensemble modeling approach has the benefit of combining the strengths of different models in predicting species distribution (Araujo and New 2007). A general linear model (GLM; R package stats [R Development Core Team 2011]) fit linear relationships (on the logit scale) between species presence and the explanatory variables and was the simplest model. A general additive model (GAM; R package gam [Hastie 2020])) allowed for fitting additional, non-linear terms to describe these relationships (Hastie and Tibshirani, 1990). The best fitting GLM model was used as the starting point for GAM analysis, where we used a stepwise approach beginning with 5 knots (points of change in the polynomials creating the curve). We ran alternate models with up to 9 knots and used AIC to choose the most parsimonious model with respect to the number of knots for each variable.

We fit 2 regression tree models for each species, a random forest (RF) and a boosted regression tree (BRT). Random forest models aggregate multiple regression trees to achieve classification of the dependent variable (R package randomForest [Liaw and Wiener 2002]). Trees in a RF model are developed from a random subset of the data such that each occurrence of the data has an equal probability of being selected for the next tree and the final aggregation incorporates each of the independently derived trees (Cutler et al. 2007). Package randomForest calculates 2 measures of variable importance: mean decrease in accuracy, a measure of how much accuracy a model would lose by eliminating individual variables, and mean decrease in Gini, a measure of the importance of an explanatory variable for estimating the outcome across all of the trees that make up the forest. A BRT model (R package gbm [Greenwell et al. 2020]) is similar to a RF model, but the data used for each tree are weighted based on the results of the previous tree such that the model continuously improves its accuracy (De'Ath 2007). Package gbm calculates a relative importance value for each explanatory variable, a measure of its relative importance in predicting the model outcome (Friedman 2001). We included all variables in the regression tree models, with 2 exceptions. We included either sagebrush or shrub (the former a subset of the latter), and we included either the shrubsteppe landscape variable (SS3km) and the CRP landscape variable (CRP3km) or the shrubsteppe/CRP landscape variable (SSCRP3km) alone (the first 2 comprising the latter), selecting the variable used in the final GLM for each species. We used default parameters in package gbm for tree complexity and learning rate in our BRT model runs. We used R package dismo (Hijmans et al. 2021) for additional model diagnostics.

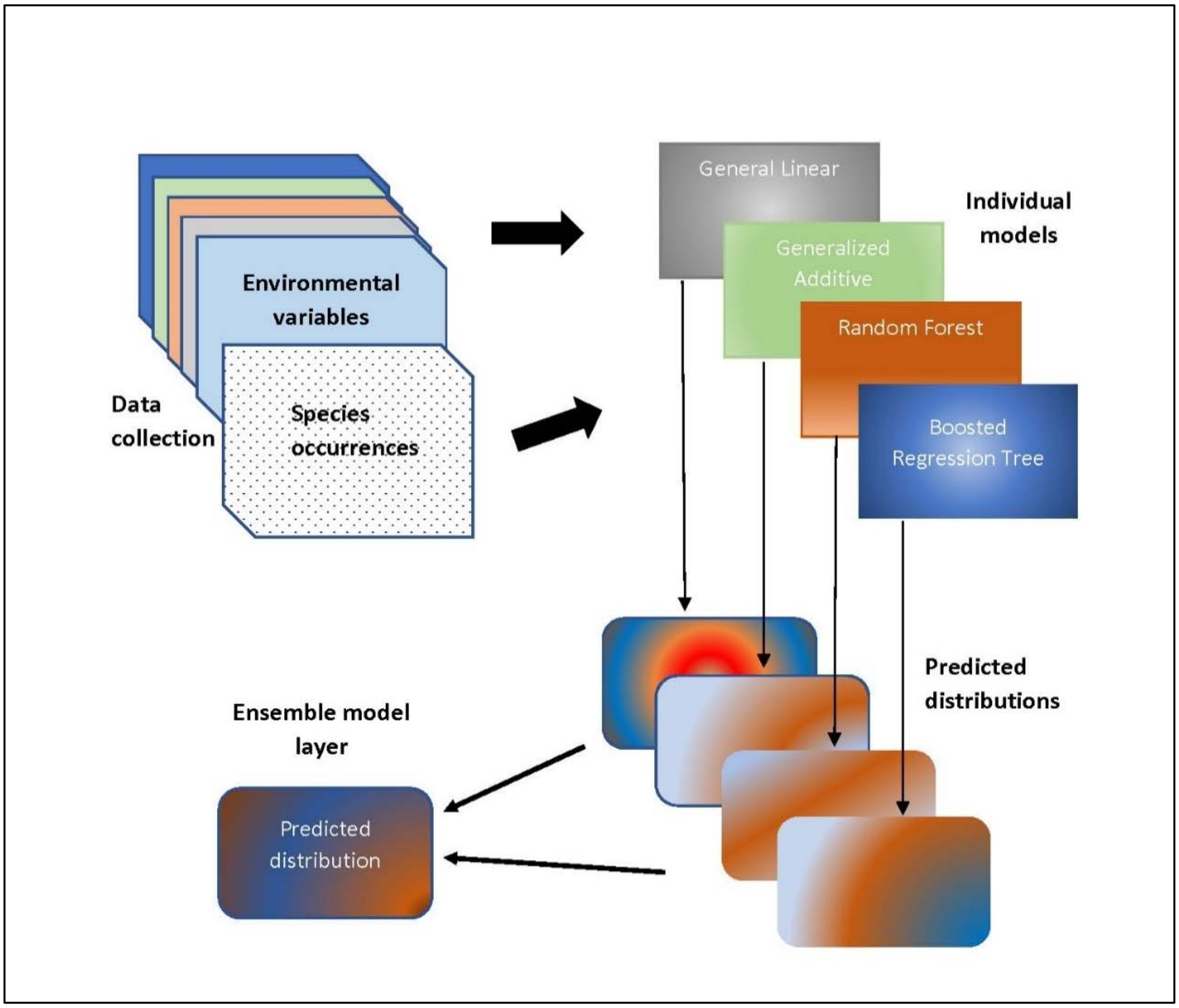


Figure 3. Process flow for combining shrubsteppe bird species occurrence data and environmental variables into models to produce predicted distributions on the landscape and combining these prediction layers to create an ensemble model layer.

Before modeling, we inspected the distribution of all explanatory variables at the 324 sites and examined correlations. Where 2 variables were highly correlated (>70%) we removed one from the pair for parametric modeling. We examined results from univariate GLMs to select variables from colinear pairs (i.e., sagebrush and shrubs; herbaceous and bare ground) and tested for quadratic effects by relating squared variables to their raw counterparts. Two of our landscape variables (SS_3km and SSCRP_3km) were highly correlated (0.92); we first entered percent SS_3km in each model and then replaced it with SSCRP_3km to examine whether CRP can benefit species that depend on extensive shrubsteppe landscapes. The combined variable SSCRP_3km represented “untilled” landscapes comprised of extant shrubsteppe and perennial grassland and shrubland provided by CRP. Where this new variable was significant it suggested that CRP worked in concert with shrubsteppe in influencing species presence, whereas where SS_3km or CRP_3km were significant it suggested different relationships between these landscape variables and species presence. We retained the landscape variables that best improved model fit.

For the parametric models (GLM and GAM) we used univariate models to screen variables for inclusion in the main models. Because our goal was predictive modeling of species distribution, we retained all variables that appeared related to occurrence of the species ($P \leq 0.1$). We also examined 2-way interaction terms for the parametric models and included them in the final models if they reduced AIC. We used the area under the receiver operator curve (AUC) as a measure of predictive performance for individual models (Wilson et al. 2005).

Developing the SDMs

The ensemble approach required fitting each final model and storing the predicted occurrence as a probability density function, along with the model-specific cut point. We selected cut point values to maximize correct classification within the data set used to build the model (Wilson et al. 2005, Freeman and Moisen 2008). A final prediction layer was then calculated as the unweighted mean of the 4 models (Araujo and New 2007). These final prediction layers mapped the probability of the species occurring in each pixel across the landscape, with values ranging from 0.05 to 0.95. We used the maximum value of the agreement Kappa statistic (MaxKappa) and the percent of observations correctly classified (PCC) to aid in selecting appropriate cut points when converting continuous probability layers to binary layers (Wilson et al. 2005, Freeman and Moisen 2008). Predicted distribution layers were limited to the spatial extent of the NLCD modeled vegetation layers and were generated at a pixel size of 180x180m.

To make these bird distribution data more readily useful for conservation, we developed model concordance layers for each species. These concordance layers attribute each pixel with the number of models that predicted presence using the individual cut points derived for each model. In the concordance layer, each pixel had a value ranging from 0 to 4, with higher values representing increasing evidence that the species might be present as estimated by the models. Conversely, a value of 0 provides the greatest evidence that the species might be absent, with higher values representing decreasing evidence. To improve the utility of the concordance layers for conservation planning and for use in further modeling efforts, we ran a smoothing algorithm (SimplifyRaster tool in ArcToolbox Ver. 10.2.6) to remove isolated pixels. This process recoded single pixels and small groups of ≤6 contiguous pixels, reducing small patches of predicted occurrences in a neighborhood of absences and vice versa. These small patches likely represented “noise” in the system, rather than small fragments of habitat, and were recoded to the value of the adjacent pixels.

We combined concurrence maps for the 3 species to illustrate areas where multiple species were predicted to occur. To achieve this, the concurrence layers for each of the species were stacked and each pixel was attributed with the number of non-zero values, resulting in summary values ranging from 0 to 3. Although any of the concurrence layers (1 to 4 model agreement) might be used for such an analysis, we selected the 2-model layer for each species to use in our example.

RESULTS

Survey results

Audubon chapters from across the state participated in the surveys, with 7 eastern Washington chapters taking the lead in establishing survey sites, training surveyors, coordinating surveys, and entering and proofing data. Over the 6 years of the project over 285 volunteers (mostly Audubon members) participated in the surveys, completing 987 individual surveys and logging over 14,000 volunteer hours.

Surveyors visited 324 sites (283 on public land and 41 on private land) between 2014 and 2019 (Fig. 4). Surveys documented 17,100 observations of 187 species, including 1497 observations of the 3 focal species. Sagebrush sparrows were detected on ≥1 survey at 81 sites (Fig. 5); on all 3 surveys at 21 sites; on 2 surveys at 19 sites; and on 1 survey at 41 sites. Sage thrashers were detected on ≥1 survey at 100 sites (Fig. 5); on all 3 surveys at 26 sites; on 2 surveys at 23 sites; and on 1 survey at 51 sites. Brewer’s sparrows were detected on ≥1 survey on 176 sites (Fig. 5); on all 3 surveys at 66 sites; on 2 surveys at 43 sites; and on 1 survey at 67 sites. Detection rates for the 3 focal species averaged 0.87 (Range 0.83-0.93) for point counts and 0.90 (0.87-0.95) for point count and traveling counts combined.

Models of habitat association

The fractional vegetation layers excluded areas that were extensive cropland and so missed some fragments of shrubsteppe in our study area. Twenty of our surveyed sites (15 public lands sites and 5 private) fell in these excluded areas and so were removed from the modeling dataset (Fig. A1), reducing the sample size to 304. Because of this limitation in the vegetation cover datasets, our species distribution maps likely underrepresent potential habitat in areas where shrubsteppe occurs only as small fragments within a cropland matrix.

Sagebrush Sparrow—The GLM for sagebrush sparrow included 5 variables that were significant in univariate models (Table A1; variable names use in models are defined in Table 2). We replaced the shrubsteppe landscape variable (SS_3km) with the combined shrubsteppe-CRP landscape variable (SSCRP_3km) because it improved model fit (Delta AIC = -5.1). We also included the 1 significant interaction (SSCRP_3km x Bare ground). We used Sagebrush in place of Shrub in the final model because

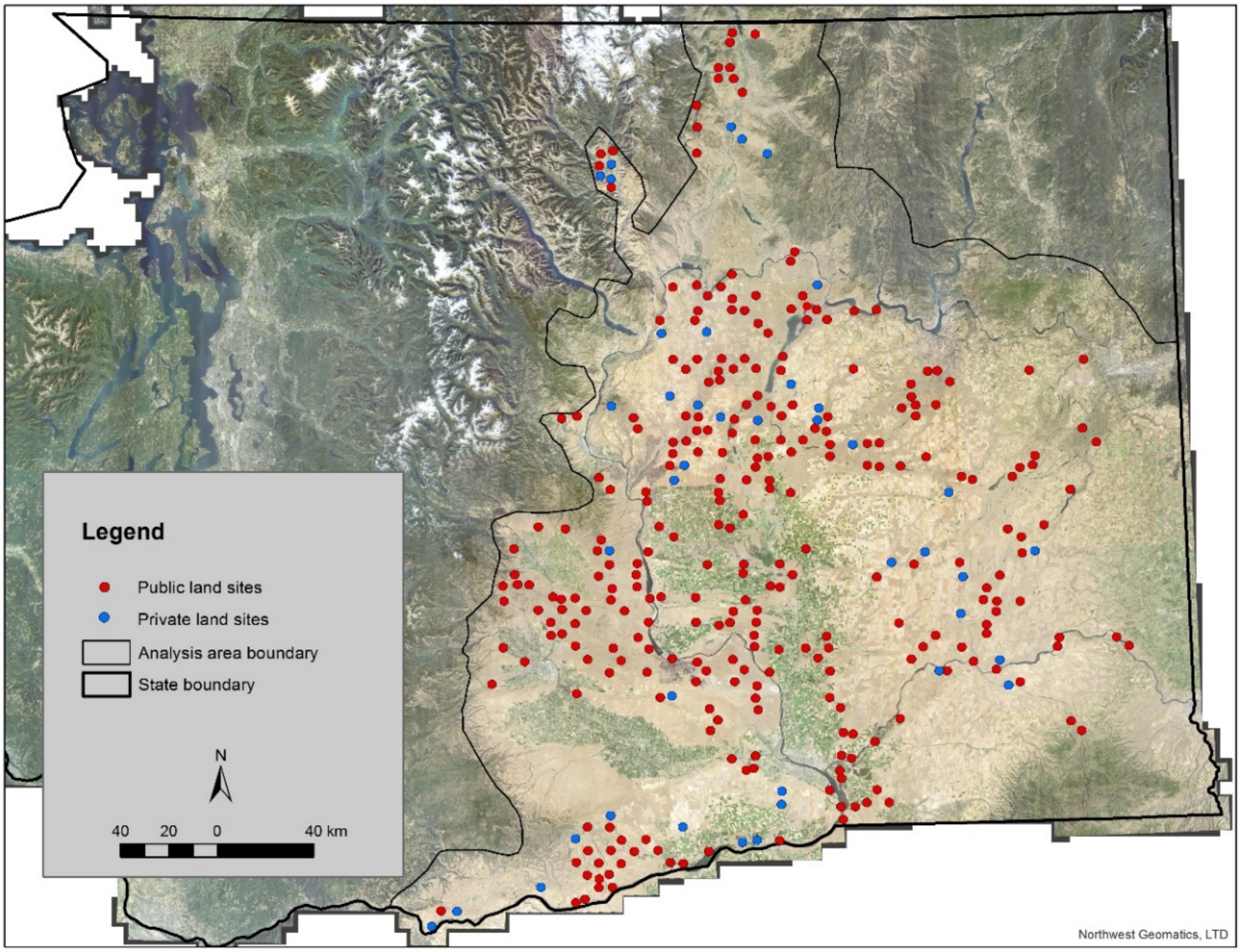


Figure 4. Location of sites surveyed for shrubsteppe birds in eastern Washington, 2014-2019.

of the close association between sagebrush sparrow and sagebrush species (see references in table 2); also, Sagebrush had a lower AIC value (4.4 units) than Shrub in univariate models. The final model had an AUC of 0.84 and we selected a cut point of 0.35 based on maximizing Kappa (0.48) and PCC (0.80). The observed prevalence (proportion of sites where the species was detected) of sagebrush sparrows at sites used in the models was 0.25.

We used the final model from the GLM as the starting model for the GAM analysis. The stepwise GAM retained all 5 variables, identifying Sagebrush and SSCRP_3km as a linear effects and Bare ground, Litter, and Longitude as non-linear effects (Table A2; see Fig. A2 for variable plots). Parameter estimates for the 2 linear effects (Sagebrush and SSCRP_3km) both were significant at α = 0.05 (Table A3). The final model had an AUC of 0.89 and we selected a cut point of 0.35 based on maximizing Kappa (0.54) and PCC (0.81).

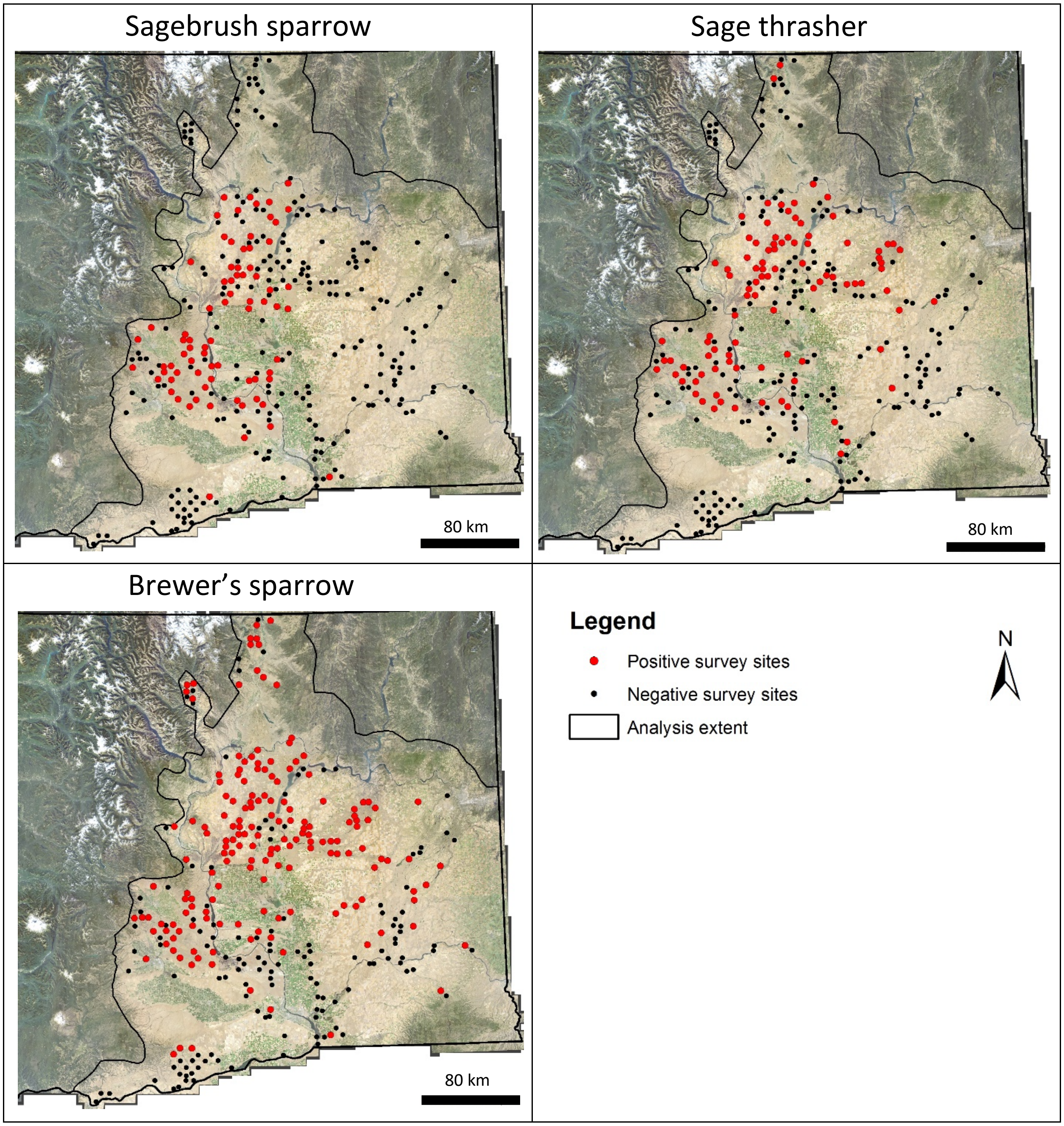


Figure 5. Results of surveys for sagebrush obligate passerines at 324 sites in eastern Washington, 2014-2019.

We included all variables in the Random Forest Model (RFM) for sagebrush sparrow except Shrub, SS_3km, and CRP_3km. The model converged and the AUC value was 0.82. The longitude variable resulted in a step function (Fig. A3) that restricted the probability map to sites on the western half of the study area with an artificial looking straight-line border; we removed the longitude variable for the final model. Bare ground and SSCRP_3km ranked highest in variable importance as indicated both by decreased mean accuracy and mean decreased Gini statistic (Fig. A4). Influence of the strongest variables on sagebrush sparrow presence is indicated by the partial dependence plots (Fig. A5), where presence increased sharply once SSCRP_3km attained values over 25% and when cover of sagebrush exceeded 5%. We selected a cut point of 0.4 based on a maximum value for Kappa (0.38) and PCC (0.77).

We ran the Boosted Regression Tree model with all explanatory variables except Shrub, SS_3km, and CRP_3km for reasons explained previously. The model converged and the AUC value was 0.81. Similar to the RFM, the longitude variable produced a step-function that restricted the probability map to sites on the west side of the analysis area; we removed the longitude variable from the final model. Also similar to the RFM, Bare ground and SSCRP_3km ranked highest in variable importance (Table A4). We selected a cut point of 0.4 based on a high value for Kappa (0.38) and maximizing the value of PCC (0.77).

Sage Thrasher—The GLM for sage thrasher had 5 variables that were significant in univariate models (Table A5). Replacing Shrub with Sagebrush resulted in a better model (-14 AIC units), whereas there was no advantage to replacing SS_3km and CRP_3km with the combined variable SSCRP_3km (+2 AIC units). There was a significant interaction between SS_3km and Sagebrush; however, including this term did not result in a better model (+0.2 AIC Units) and we chose not to include it. The final model had an AUC of 0.82 and we selected a cut point of 0.4 based on maximizing Kappa (0.47) and PCC (0.76). The observed prevalence of sage thrashers at sites used in the models was 0.32.

The stepwise GAM retained all 5 variables from the GLM, identifying Sagebrush, SS_3km, CRP_3km, and Bare ground highly significant as a linear effects and elevation as a non-linear effect (Table A6; see Fig. A6 for variable plots). Parameter estimates for the 4 linear effects (Sagebrush, SS_3km, CRP_3km, and Bare ground) all were significant at $\alpha = 0.05$ (Table A7). The final model had an AUC of 0.83 and we selected a cut point of 0.4 based on maximizing Kappa (0.54) and PCC (0.80).

We included all variables in RFM except Shrub and SSCRP_3km. The model converged and the AUC value was 0.81. Sagebrush, Elevation, and Latitude ranked highest in variable importance as indicated both by decreased mean accuracy and mean decreased Gini statistic (Fig. A7). Influence of the strongest variables on sage thrasher presence is indicated by the partial dependence plots (Fig. A8), where presence generally increased with elevation and latitude and with increasing cover of sagebrush. We selected a cut point of 0.35 based on maximizing Kappa (0.47) and PCC (0.76).

We ran the BRT model with all explanatory variables except Shrub and SSCRP_3km. The model converged and the AUC value was 0.78. Sagebrush, Elevation, SS_3k, and Latitude ranked highest in variable importance (Table A8). We selected a cut point of 0.4 based on maximizing the value of Kappa (0.43) and PCC (0.76).

Brewer's Sparrow—The GLM for Brewer's sparrow included 8 variables that were significant in univariate models, including the quadric transformation of the Herbaceous variable (Herbsq) (Table A9). Herbsq was marginally significant in the final model and including Herbsq resulted in a better model (-6.0 AIC units). Although all 3 landscape variables were significant in univariate models, none resulted in a better final model (greater AIC values in all cases). The final model had an AUC of 0.89 and we selected a cut point of 0.55 based on maximizing Kappa (0.58) and PCC (0.79). The observed prevalence of Brewer's sparrows at sites used in the models was 0.56.

The stepwise GAM retained all 7 variables from the GLM; we did not include the quadratic term for Herbaceous, allowing the GAM to identify any non-linear relationships. The GAM identified Agsoil, Elevation, Herbaceous, and Sagebrush significant as a linear effects and Longitude and Litter as a non-linear effects (Table A10; see Fig. A9 for variable plots). The final model had an AUC of 0.90 and we selected a cut point of 0.55 based on maximizing Kappa (0.64) and PCC (0.82). Parameter estimates for the linear effects were significant (Table A11).

We included all variables in the RFM except Shrub and SSCRP_3km. The model converged and the AUC value was 0.85. Longitude, DEM, and Sagebrush ranked highest in variable importance as indicated both by decreased mean accuracy and mean decreased Gini (Fig. A10). We selected a cut point of 0.35 based on maximizing Kappa (0.56) and PCC (0.78). Partial dependence plots for variables are in Fig. A11.

We ran the BRT with all explanatory variables except Shrub and SSCRP_3km. The model converged and the AUC value was 0.85. Longitude, DEM, and Sagebrush ranked highest in variable importance (Table A12). We selected a cut point of 0.56 based on maximizing the value of Kappa (0.55) and PCC (0.78).

Ensemble Models and Distribution Maps

Sagebrush Sparrow—The ensemble model for sagebrush sparrow produced a probability layer with high values centered in the largest expanses of extant shrubsteppe on federal and state lands west of the Columbia River and extending north along Moses coulee (Fig. 6). Positive effects for sagebrush cover and proportion of the landscape in shrubsteppe and CRP (SSCRP_3km) were strong in the model set. High values for the Longitude variable in some models was reflected in the lack of any observations on our sites in the eastern portion of the study area. The concordance layers further documented this trend, with the modeled distribution collapsing to large areas of extensive shrubsteppe in the western part of the study area as models converged (Fig. 7).

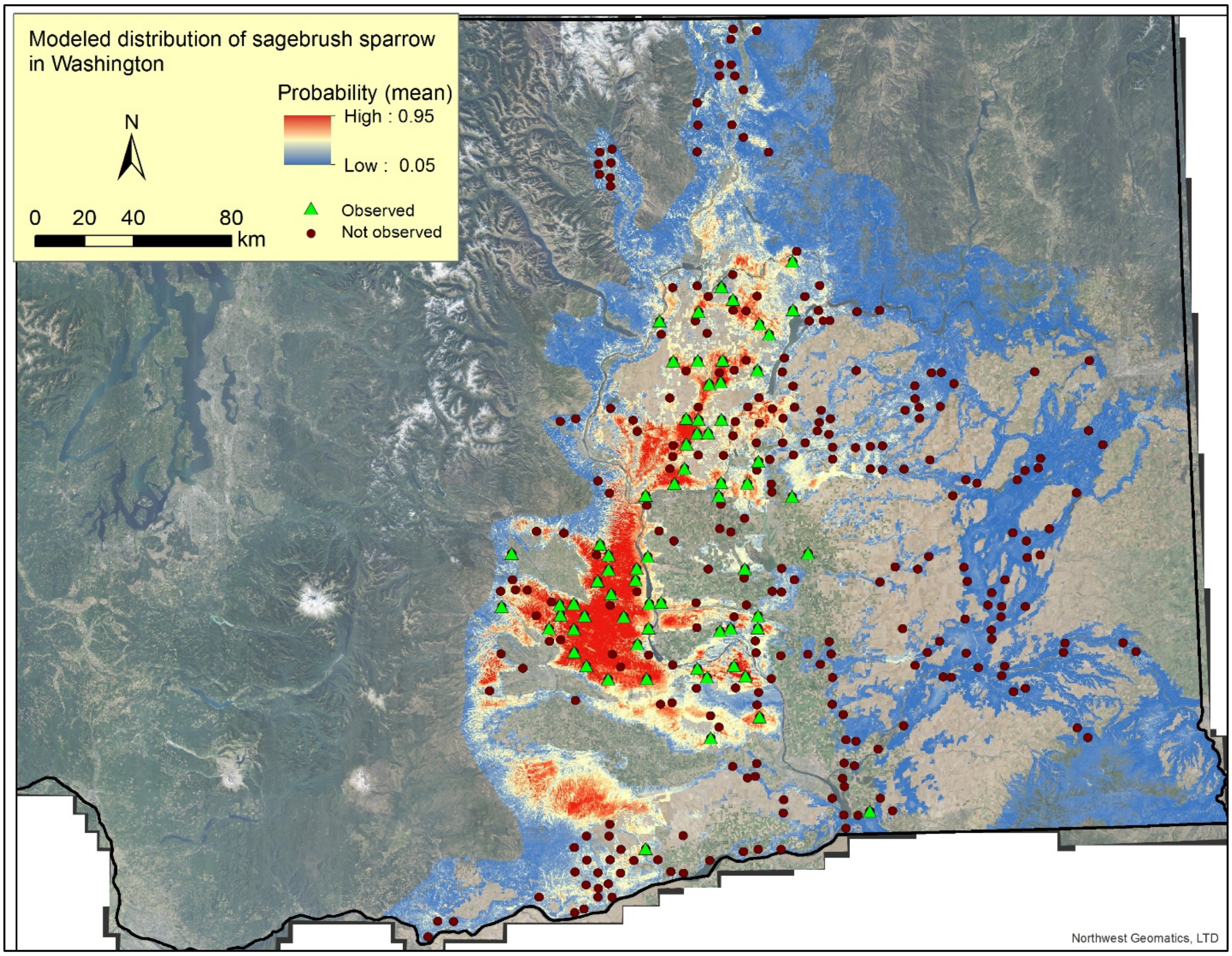


Figure 6. Modeled distribution and survey results for sagebrush sparrow in Washington State based on ensemble modeling of survey data from 304 locations, 2014-2019. Points illustrate survey locations and result (observed or not observed).

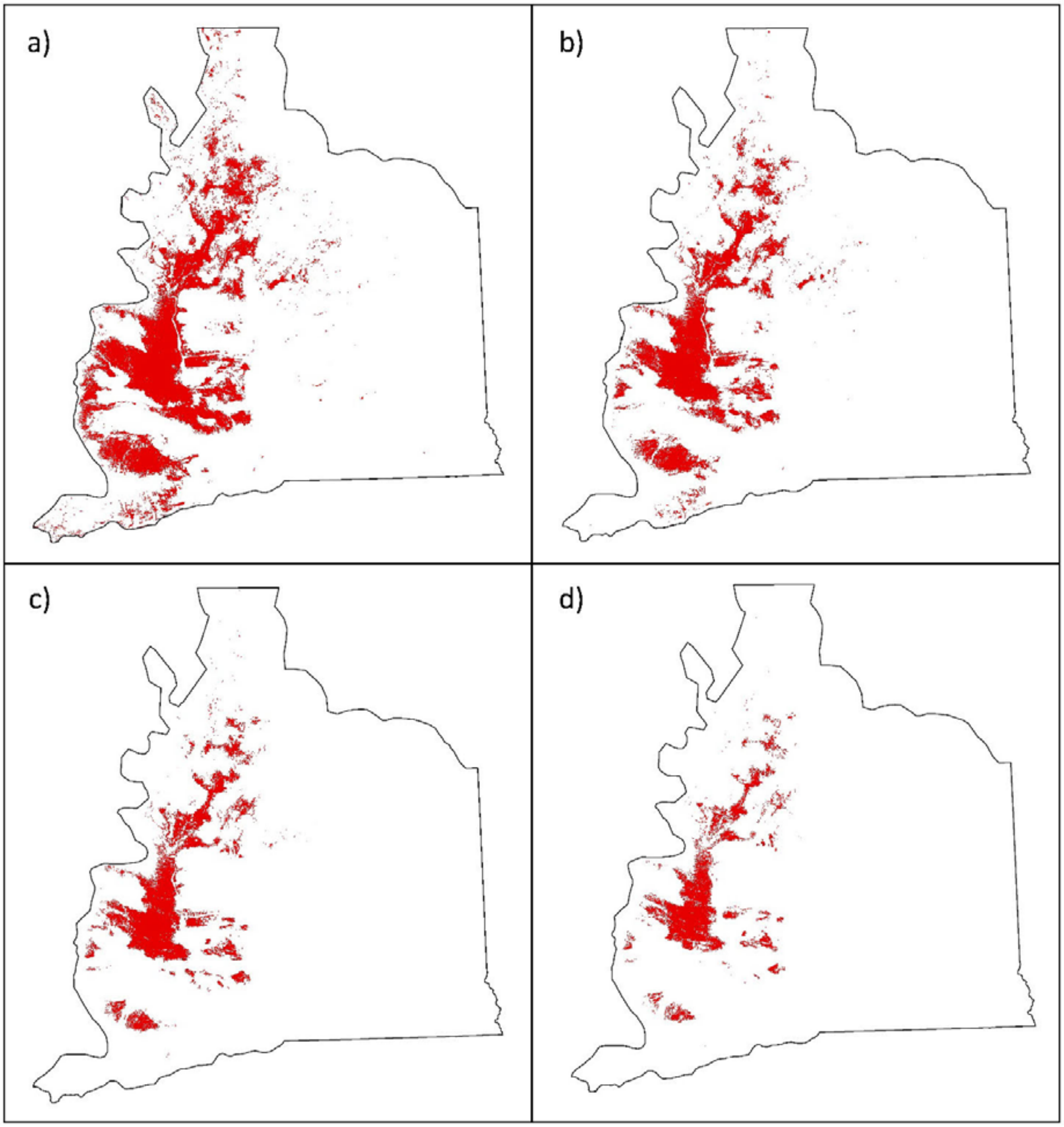


Figure 7. Modeled distribution of sagebrush sparrows within the analysis area in Washington State. Distributions are based on ensemble modeling of survey data from 304 locations, 2014-2019. Panels a) through d) illustrate predicted distribution based on agreement of from 1 to 4 models, respectively.

Sage Thrasher—The ensemble model for sage thrasher produced a probability layer with high values centered in the largest expanses of extant shrubsteppe on federal and state lands west of the Columbia River, but unlike the sagebrush sparrow, the layer predicted occurrence north into the Okanogan valley and east into the channeled scablands of Lincoln County (Fig. 8). Cover of sagebrush was the strongest predictor across all models and both landscape variables (SS_3km and CRP_3km) also were strong across models. Similar to the case with sagebrush sparrow, the concordance layers illustrate the modeled distribution collapsing to large areas of extensive shrubsteppe, but without a northing or easting trend, as models converged (Fig. 9).

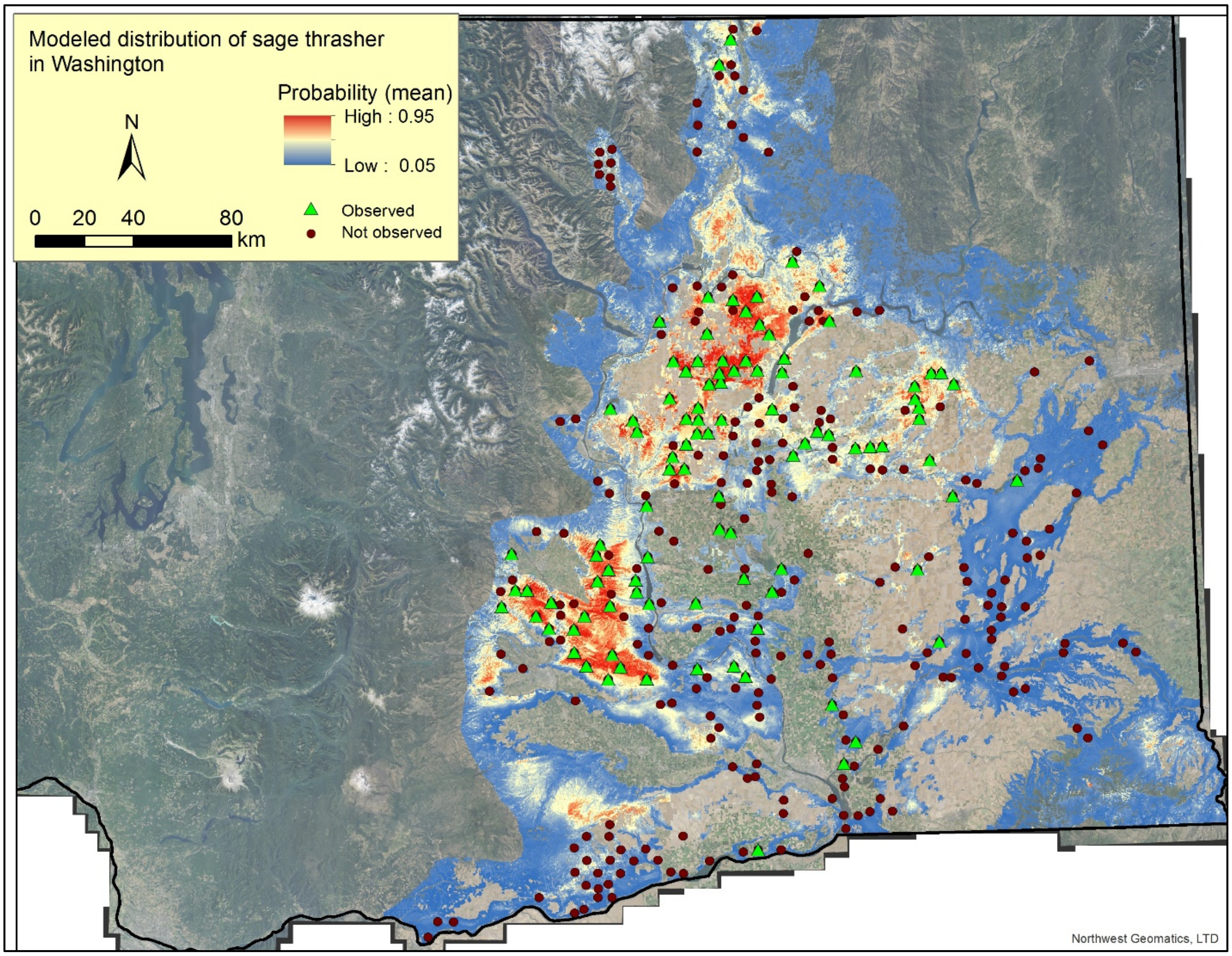


Figure 8. Modeled distribution and survey results for sage thrasher in Washington State based on ensemble modeling of survey data from 304 locations, 2014-2019. Points illustrate survey locations and result (observed or not observed).

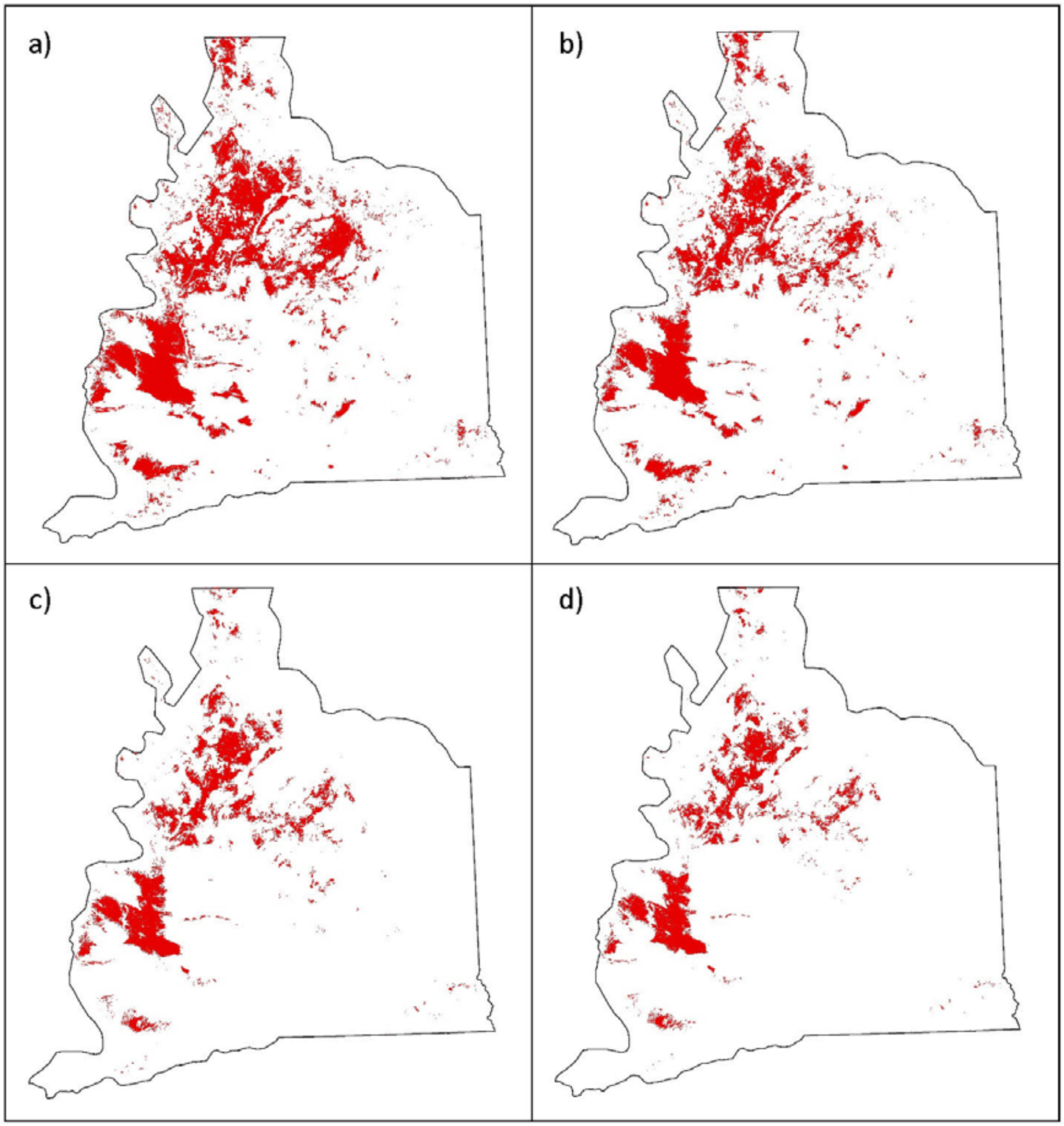


Figure 9. Modeled distribution of sage thrasher within the analysis area in Washington State. Distributions are based on ensemble modeling of survey data from 304 locations, 2014-2019. Panels a) through d) illustrate predicted distribution based on agreement of from 1 to 4 models, respectively.

Brewer's Sparrow—The ensemble model for Brewer's sparrow produced a probability layer with high values spanning central and northern areas of the Columbia Plateau and extending into the Blue Mountains ecoregion in the southeast corner of the state (Fig. 10). Strong, positive effects for the Latitude variable and sagebrush cover were common to all models and were reflected in the spatial patterns of highest probability. Unlike the other 2 sagebrush obligates, Brewer's sparrows were not strongly affected by landscape variables. The concordance layers further documented the trend in more northerly occurrence, with the modeled distribution collapsing to the north and away from the Blue Mountains ecoregion as models converged (Fig. 11).

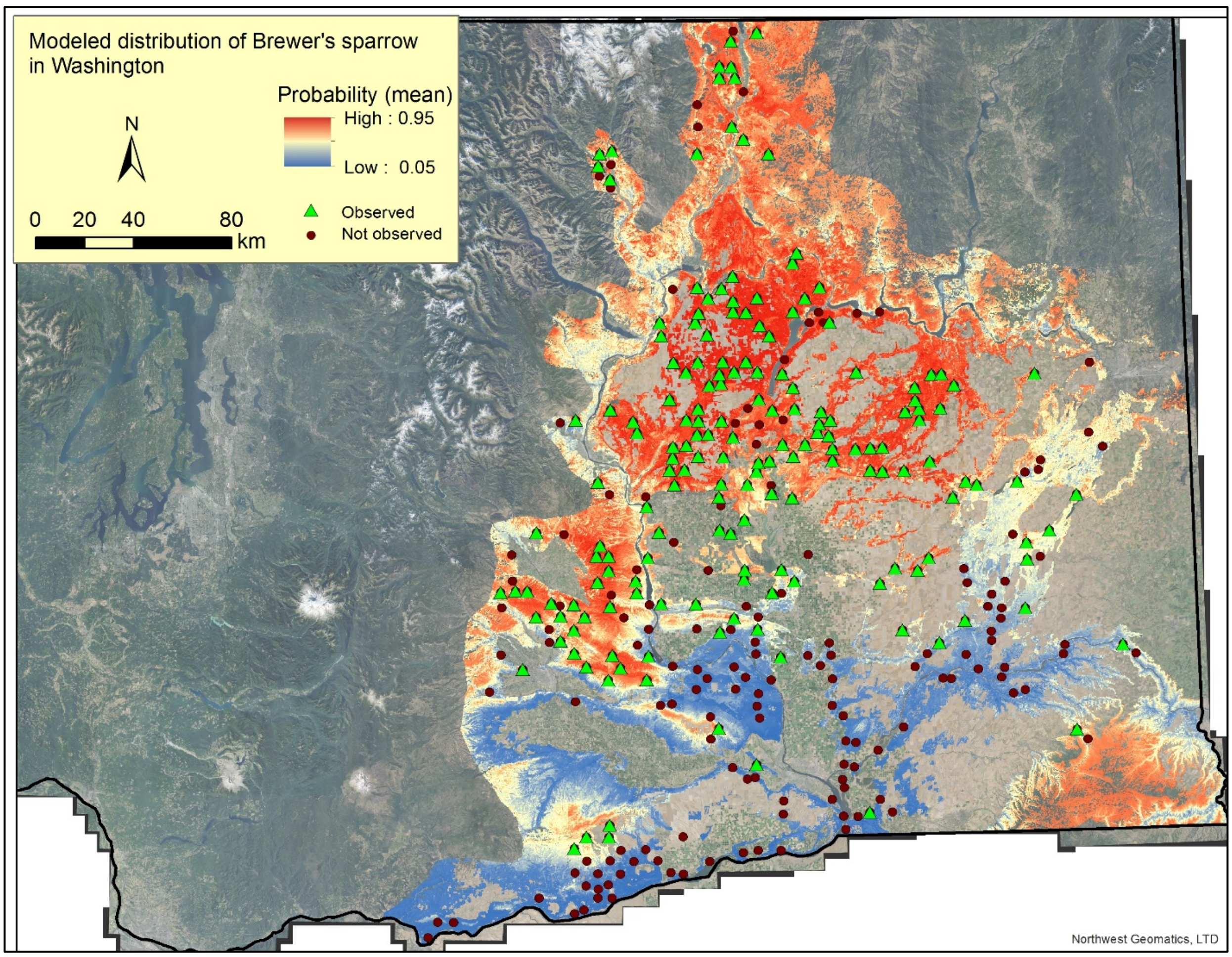


Figure 10. Modeled distribution and survey results for Brewer's sparrow in Washington State based on ensemble modeling of survey data from 304 locations, 2014-2019. Points illustrate survey locations and result (observed or not observed).

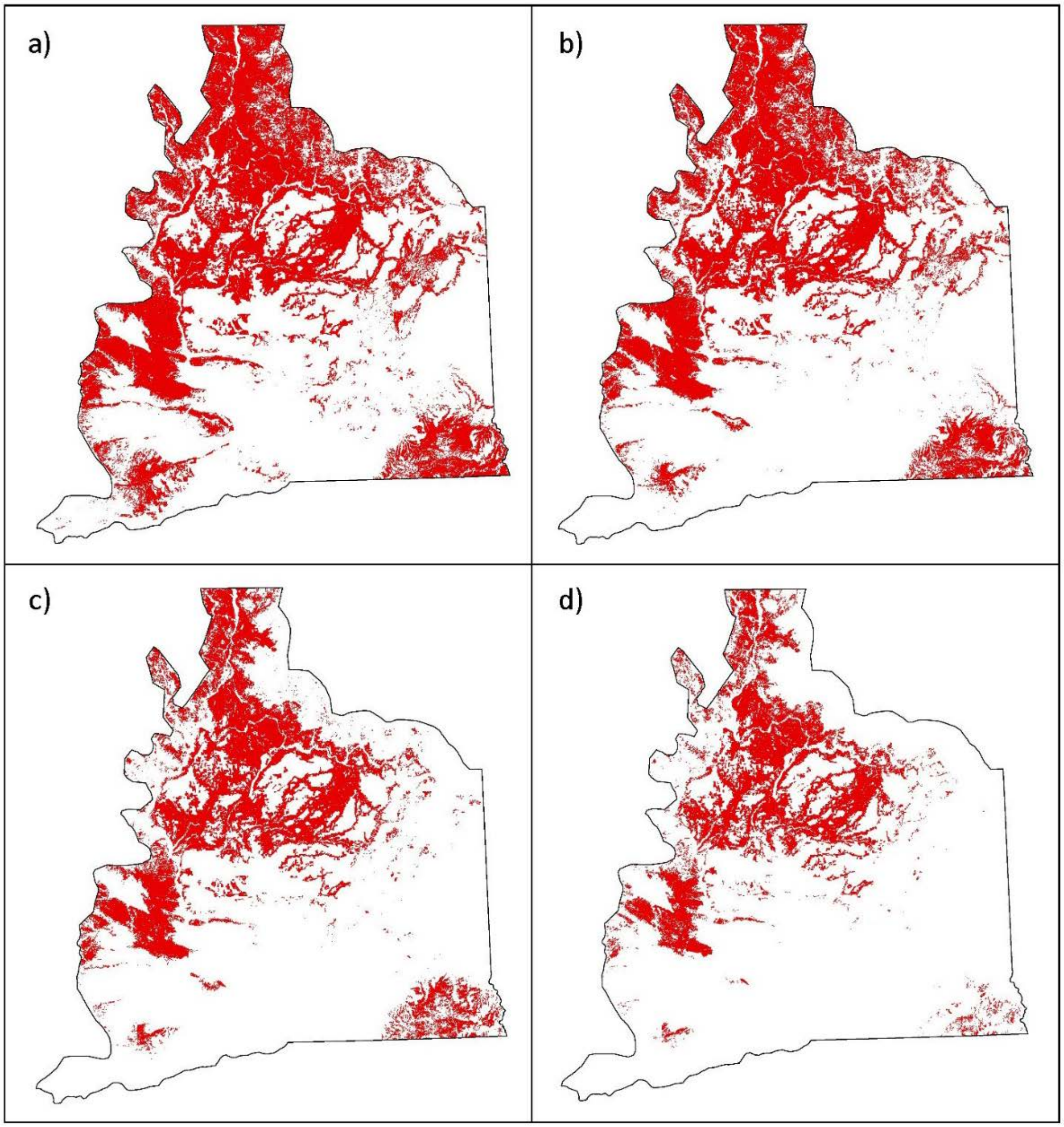


Figure 11. Modeled distribution of Brewer's sparrows within the analysis area in Washington State. Distributions are based on ensemble modeling of survey data from 304 locations, 2014-2019. Panels a) through d) illustrate predicted distribution based on agreement of from 1 to 4 models, respectively.

Multi-species Distribution Layers

Combining model concurrence layers for the 3 species illustrates areas in Washington where multiple species might be expected to be present (Fig. 12). Sagebrush sparrow, with the most restricted distribution predicted by our models, drives the 3-species category, limiting it to the westernmost part of the analysis area. Areas of the state predicted to have none of the 3 modeled species primarily occur along the snake river and north through the channeled scablands. These areas also were predicted to have low or no cover of sagebrush by the NCLD fractional data layer. As with the single-species distribution layers, it is important to keep in mind that the binary layers that went into creating this multi-species layer each are based on a probability density function that attributed each pixel from high to low probability of occurrence. Areas attributed with a zero predicted occurrence in binary layers actually equate to a low probability of occurrence relative to other areas (i.e., the species may still be present, but the predicted probability of that occurring is low).

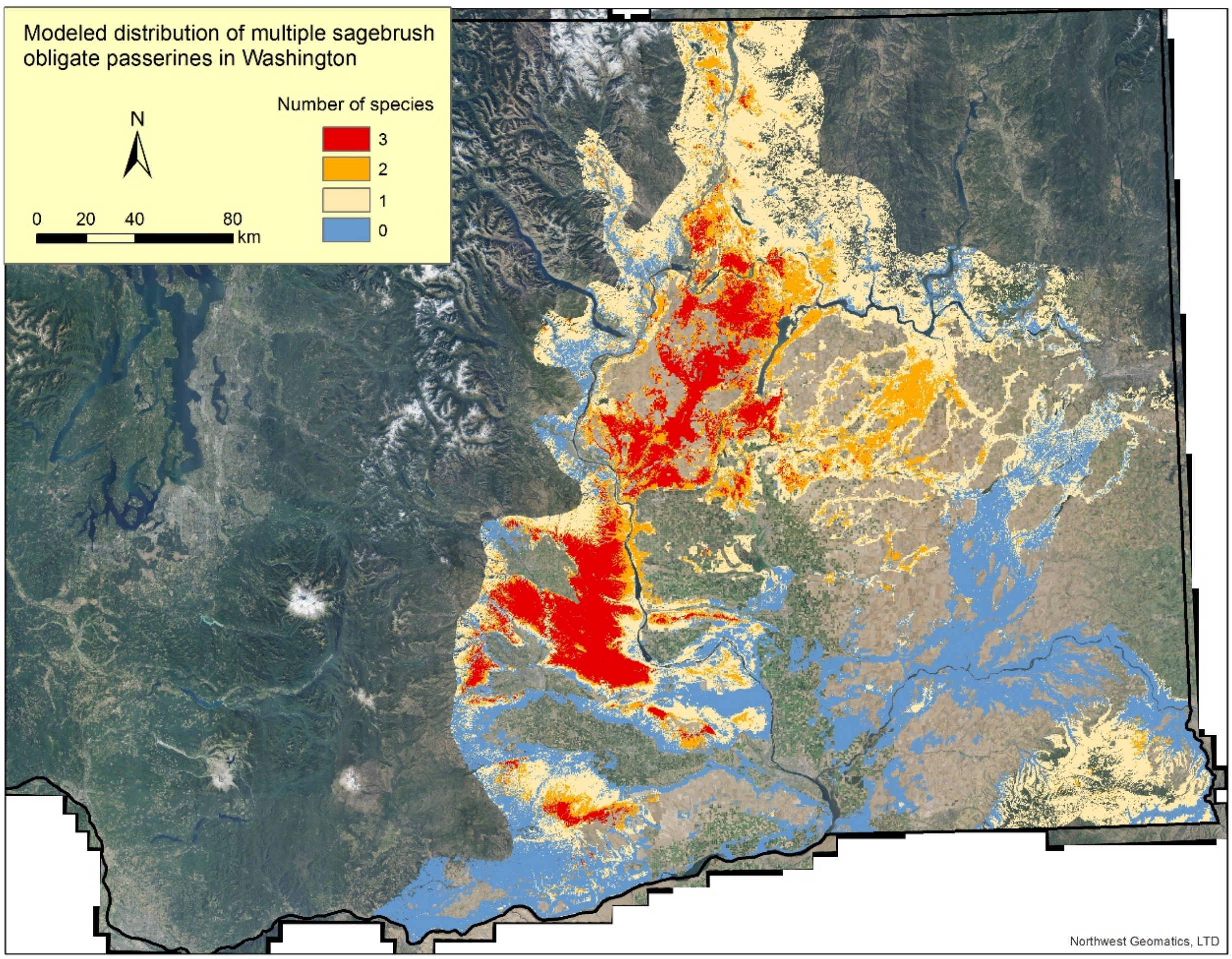


Figure 12. Modeled distribution of sagebrush sparrows, sage thrashers, and Brewer's sparrows within the analysis area in Washington State. Layer depicts the number of species (0-3) predicted to occur in each pixel based on ensemble modeling of survey data from 304 locations, 2014-2019.

DISCUSSION

Habitat Associations from the Models

The results from our models provide additional insight into species-habitat relationships as well as providing the basis for the predicted distribution maps discussed below. The most significant variables found to influence species presence in our study aligned well with results from earlier research based on field measurements at occupied and unoccupied sites. Sagebrush sparrows generally occur only where there is sagebrush, and probability of occurrence has been found to increase with sagebrush cover (Wiens and Rotenberry 1981, Knick and Rottenbery 1995, Earnst and Holmes 2012, Vander Haegen et al. 2000, 2015). Bare ground for foraging also is a characteristic of sagebrush sparrow habitat (Rotenberry and Wiens 1980), so the high ranking of this variable in all models was not surprising. The shape of the curve for Bare ground in the GAM for sagebrush sparrow (Fig. A2) suggests that the relationship is not linear for the sites that we sampled, but instead decreased at low and high values and increased across intermediate values.

Sagebrush sparrows are particularly affected by landscape fragmentation, and this pattern established in earlier research was repeated in our findings. Studies in Idaho (Knick and Rottenbery 1995) and in eastern Washington (Vander Haegen et al. 2000) found that sagebrush sparrows preferred intact shrubsteppe landscapes; moreover, this species may experience greater reproductive success in these areas compared to more fragmented landscapes (Vander Haegen 2007). An important addition from the present study is the strong relationship identified between sagebrush sparrow presence and the degree of surrounding landscape in shrubsteppe and CRP across all 4 of our models. By combining the 2 landscape variables in our analysis (SS_3km and CRP_3km) we examined the influence of an extensive, untilled landscape on presence of sagebrush sparrows, and this proved to have a stronger effect than including shrubsteppe and CRP variables separately. While it has been hypothesized in the past that the vegetation structure and perennial nature of CRP fields may supplement extant shrubsteppe in creating untilled landscapes suitable for area sensitive species (Vander Haegen et al. 2000, 2015), findings from our study offer support for that relationship with sagebrush sparrows.

The strongest variables in our sage thrasher models also aligned well with results from previous work. Percent cover of shrubs (Vander Haegen et al. 2000), or more specifically sagebrush (Knick and Rotenberry 1995), was positively associated with sage thrashers as was greater percentage of open ground (Rotenberry and Wiens 1980). Sage thrashers are influenced by landscape context, with research in Idaho reporting a positive association between thrashers and the percentage of shrubsteppe in the surrounding landscape (Knick and Rottenbery 1995). Earlier work in Washington reported the opposite association (Vander Haegen et al. 2000), and indeed sage thrashers have been found nesting in small shrubsteppe fragments in Washington (Vander Haegen 2007). The present study supports the positive association reported from the earlier work in Idaho, although inspection of the partial dependence plots suggests a more complex relationship, with an initial decrease in occurrence of thrashers as shrubsteppe in the landscape increases followed by an increasing trend (Fig. A8). Unlike sagebrush sparrows, sage thrashers in Washington will establish territories and nest in landscapes fragmented by agriculture, although at the cost of reduced nesting success (Vander Haegen 2007). The positive association between sage thrasher presence and both the percent of the landscape in shrubsteppe and in CRP suggests a preference for untilled landscapes, while the different shapes of the curves for these 2 variables (Fig. A8) indicate subtle differences in the relationship.

Unlike sagebrush sparrows, Brewer's sparrows can be found in small shrubsteppe fragments as well as in extensive areas of unbroken shrubsteppe. Similar to previous studies that have examined landscape effects on shrubsteppe birds (Knick and Rotenberry 1995, Vander Haegen et al. 2000, 2015) we found no strong association between presence of Brewer's sparrows and the percentage of shrubsteppe in the surrounding landscape. Brewer's sparrows will nest in small shrubsteppe fragments, but as with sage thrashers, at the cost of reduced nesting success (Vander Haegen 2007). The strongest habitat association for Brewer's sparrow found in our study was with percent cover of sagebrush, a relationship that has been well documented by previous research (Larson and Bock 1984, Knick and Rotenberry 1995, Vander Haegen et al. 2000, 2015). Like the sagebrush sparrow and sage thrasher, Brewer's sparrows in Washington most often nest in sagebrush plants and forage in sagebrush and in the herbaceous layer below (Rotenberry et al. 2020). While there was evidence of an association between Brewer's sparrow occurrence and percent cover of herbaceous plants and cover of litter, the relationships were not strong.

Species Distribution Maps

Our SDMs based on presence/absence data collected by community science illustrate both the extent of these species' predicted occurrence in Washington and the fragmented nature of the landscapes where they occur. For our study of shrubsteppe birds, both the survey design and the availability of suitable habitat layers made SDM an appropriate and useful approach (Rondinini et al. 2006). By systematically selecting survey sites over a large geographic area and designing surveys to maximize detection of our focal species we enhanced both the confidence of our presence/absence data (Mackenzie and Royle 2005) and the suitability of those data for building SDMs (Sofaer et al. 2019). Importantly, the availability of recent spatial vegetation layers made it possible for us to employ reasonable approximations of important habitat components in our spatial models of bird distribution (Elith and Leathwick 2009). Previous distribution layers for these species from the Washington Bird Atlas project were based on mapped vegetation zones and a landcover layer developed from Landsat Thematic Mapper data (Smith et al. 1997). These earlier maps lack the spatial refinement made possible by using SDMs informed by fine-scale vegetation layers and landscape metrics, and likely overestimated the potential distribution of these species.

Data from SDMs can be presented in multiple ways to meet specific objectives (Araujo and New 2007). For our study, we chose 2 approaches: a probability density function that was an average of all 4 models, and a consensus or frequency approach that considered the results of individual models (Araujo and New 2007). Both began with probabilistic results from the individual models (returning values between 0.05 and 0.95 for each pixel), with estimates spanning the geographic area defined by the available habitat layers. The final probability layers for each species were based on an ensemble of 4 models, where the value for each pixel was an unweighted mean of the predicted occurrence across these models. We chose to not weight individual models given that they all performed well and had similar fit to the data (Araujo and New 2007). While the probability density layers are an effective way to view and display the data and can be used in conservation efforts (Guisan et al. 2013), converting the pixel probabilities to binary values is the more commonly used approach for use in conservation planning (Wilson et al. 2005, Rondinini et al. 2006).

Concurrence Maps

Creating binary data layers of species distribution allowed us to more clearly define areas of potential importance for use in conservation planning. By using an ensemble approach in our modeling, we were able to leverage the concurrence of individual models agreeing on the likelihood of species being present as an index to our certainty in the predictions. We created 4 raster layers for each species, each layer representing concurrence of from 1 to 4 models. More specifically, layers indicating 1 model agreement depict all pixels where at least 1 of the 4 models predicted presence of that species. These layers are the least conservative of the set and represent the broadest depiction of potential distribution from the modeled data. In contrast, layers indicating agreement by 4 models depict only pixels where all models predicted presence of that species. These layers are the most conservative of the set and represent the narrowest depiction of potential distribution from the modeled data. The different model layers, ranging from 1 to 4, can be thought of as representing increasing evidence of presence (or, conversely, decreasing evidence of absence) for each species as estimated by the models (Fig. 13). Selecting which of the 4 concurrence layers to use will depend on the application and can be thought about as the relative need for avoiding errors of commission (false positives) or omission (false negatives) (Loiselle et al. 2003). The following examples may make this more clear.

Example 1. Surveying for rare species on the landscape can be prohibitively expensive both in time and funds; using SDM layers to focus survey efforts in areas more likely to have suitable habitat can greatly improve the efficacy of these surveys (Edwards et al. 2005). As an extension, SDMs can be useful for identifying areas that warrant being surveyed for a species of concern when planning, or prior to approving, a management action. Focusing the areas for survey in these examples might best be achieved with the 1 or 2 model layers (Fig. 13). While these less conservative layers are more likely to include areas less suitable to the species, they allow managers greater confidence that potential effects on the species of concern were adequately considered (low omission error).

Example 2. As certainty in predicting species presence becomes more important in conservation planning, layers with increased model agreement are more appropriate. An example here might be identifying sites for habitat reserves, where managers desire high confidence that the species of concern is being afforded protection (low commission error). Under this scenario, identifying potential sites for conservation might best be achieved with the 3 or 4 model layers (Fig. 13). Using the less conservative 1 or 2 model layers for reserve protection runs the risk of conserving land that is less likely to provide adequate habitat for the species (Loiselle et al. 2003). Combining model concurrence layers for multiple species can be an effective way to illustrate areas with high conservation potential; the same considerations apply regarding selection of the most appropriate concurrence maps to use in building a multi-species layer.

# Models in agreement	Evidence/ certainty	Benefits	Potential applications
1	Less	Less likely to exclude actual habitat or occupied areas	• General distribution mapping • Planning surveys for potential management actions (e.g., wind energy projects)
2			
3			• Delineating high value areas for conservation. • Assessing quantity and distribution of protected habitat • Estimating changing habitat availability over time
4	More	Less likely to include non-habitat and unoccupied areas	

Figure 13. Guide to selecting model concurrence layers of shrubsteppe bird species distribution for application to conservation and management goals in Washington.

CONCLUSIONS

Rigorous surveys conducted over 6 years using community science provided data suitable to develop SDMs for at-risk bird species in Washington. The models identified site and landscape variables associated with presence of each species, some that confirmed associations found in previous research and one that provided nuance to an already recognized relationship. While both sagebrush sparrows and sage thrashers have been found to prefer extensive shrubsteppe landscapes, our data suggest that having CRP in the landscape in place of tilled agricultural fields may enhance use of some fragmented landscapes by these species.

The species distribution layers fill a critical information gap that can aid in the conservation of these shrubsteppe obligate passerines by informing land use planning, selection of conservation reserve areas, and other management actions that would benefit from detailed prediction of were species might be expected to occur. Each set of 4 model concurrence layers provides flexibility when predicting species presence on the landscape, ranging from the broadest depiction of potential distribution that minimizes omission error (1 model layers) to the most conservative depiction that minimizes commission error (4 model layers). Our results suggest that community science, particularly when coupled with proper study design and a motivated, well-run volunteer organization, can be an effective tool for informing pressing conservation needs.

LITERATURE CITED

Araújo, M. B., and M. New. 2007. Ensemble forecasting of species distributions. Trends in Ecology & Evolution 22:42-47.

Braun, C. E., M. F. Baker, R. L. Eng, J. S. Gashwiler, and M. H. Schroeder. 1976. Conservation committee report on effects of alteration of sagebrush communities on the associated avifauna. Wilson Bulletin 88:165-171.

Cutler, D. R., T. C. Edwards, Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

De'Ath, G. 2007. Boosted trees for ecological modeling and prediction. Ecology 88:243-251.

Dobler, F. C., J. Eby, C. Perry, S. Richardson, and M. Vander Haegen. 1996. Status of Washington's shrub-steppe ecosystem: extent, ownership, and wildlife/vegetation relationships. Research Report, Washington Department of Fish and Wildlife, Olympia.

Earnst, S. L., and A. L. Holmes. 2012. Bird-habitat relationships in interior Columbia Basin shrubsteppe. Condor 114:15-29.

Edwards Jr, T. C., D. R. Cutler, N. E. Zimmermann, L. Geiser, and J. Alegria. 2005. Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86:1081-1090.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677-697.

Freeman, E. A., and G. G. Moisen. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecological Modelling 217:48-58.

Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics 28:1189-1232.

Greenwell, B., B. Boehmk, J. Cunningham, and G. Developers. 2020. Gbm: Generalized boosted regression models. R Package version 2.1.8.

Guisan, A., R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. T. Tulloch, T. J. Regan, L. Brotons, E. McDonald-Madden, C. Mantyka-Pringle, T. G. Martin, J. R. Rhodes, R. Maggini, S. A. Setterfield, J. Elith, M. W. Schwartz, B. A. Wintle, O. Broennimann, M. Austin, S. Ferrier, M. R. Kearney, H. P. Possingham, and Y. M. Buckley. 2013. Predicting species distributions for conservation decisions. Ecology Letters 16:1424-1435.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147-186.

Hastie, T. 2020. Gam: Generalized additive models. R package version 1.20.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman & Hall, Boca Raton, Florida, USA.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2021. Dismo: Species distribution modeling. R Package version 1.3.5.

Jacobson, J. E., and M. C. Snyder. 2000. Shrubsteppe mapping of eastern Washington using Landsat satellite thematic mapper data. Research Report. Washington Department of Fish and Wildlife, Olympia.

Knick, S. T., D. S. Dobkin, J. T. Rotenberry, M. A. Schroeder, W. M. Vander Haegen, and I. C. Van Riper. 2003. Teetering on the edge or too late? Conservation and research issues for avifauna of sagebrush habitats. Condor 105:611-634.

Knick, S. T., and J. T. Rotenberry. 1995. Landscape characteristics of fragmented shrubsteppe habitats and breeding passerine birds. Conservation Biology 9:1059-1071.

Larson, D. L., and C. E. Bock. 1984. Determining avian habitat preference by bird-centered vegetation sampling. Pages 37-43 *in* Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison, USA.

Liaw, A., and M. Wiener. 2002. Classification and regression by random forest. R News 2:18-22.

Loiselle, B. A., C. A. Howell, C. H. Graham, J. M. Goerck, T. Brooks, K. G. Smith, and P. H. Williams. 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology 17:1591-1600.

MacKenzie, D. I., and J. A. Royle. 2005. Designing occupancy studies: General advice and allocating survey effort. Journal of Applied Ecology 42:1105-1114.

Millikin, R. L., R. Joy, J. Komaromi, M. Harrison, N. Mahony, and W. M. Vander Haegen. 2020. Critical habitat identification of peripheral sage thrashers under climate change. Conservation Science and Practice 2:e290

Pollock, K. H., J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, and J. R. Sauer. 2002. Large scale wildlife monitoring studies: Statistical methods for design and analysis. Environmetrics 13:105-119.

R Development Core Team. 2011. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Rigge, M., C. Homer, L. Cleeves, D. K. Meyer, B. Bunde, H. Shi, G. Xian, S. Schell, and M. Bobo. 2020. Quantifying western U.S. Rangelands as fractional components with multi-resolution remote sensing and in situ data. Remote Sensing 12, 412; doi:10.3390/rs12030412.

Rondinini, C., K. A. Wilson, L. Boitani, H. Grantham, and H. P. Possingham. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecology Letters 9:1136-1145.

Rotenberry, J. T., M. A. Patten, and K. L. Preston (2020). Brewer's Sparrow (*Spizella breweri*), version 1.0 *In* Birds of the World (A. F. Poole and F. B. Gill, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.brespa.01

Rotenberry, J. T., and J. A. Wiens. 1980. Habitat structure, patchiness, and avian communities in north American steppe vegetation: A multivariate analysis. Ecology 61:1228-1250.

Schroeder, M. A., and W. M. Vander Haegen. 2011. Response of greater sage-grouse to the conservation reserve program in Washington state. Studies in Avian Biology 38:517-529.

Sofaer, H. R., C. S. Jarnevich, I. S. Pearse, R. L. Smyth, S. Auer, G. L. Cook, T. C. Edwards, Jr., G. F. Guala, T. G. Howard, J. T. Morisette, and H. Hamilton. 2019. Development and delivery of species distribution models to inform decision-making. BioScience 69:544-557.

Smith, M. R., P. W. Mattocks, Jr., and K. M. Cassidy. 1997. Breeding birds of Washington state. Vol. 4 in Washington State Gap Analysis--Final Report (K. M. Cassidy, C. E. Grue, M. R. Smith, and K. M. Dvornich, eds.). Seattle Audubon Society Publications in Zoology No. 1, Seattle.

Sullivan, B. L., J. L. Aycrigg, J. H. Barry, R. E. Bonney, N. Bruns, C. B. Cooper, T. Damoulas, A. A. Dhondt, T. Dietterich, A. Farnsworth, D. Fink, J. W. Fitzpatrick, T. Fredericks, J. Gerbracht, C. Gomes, W. M. Hochachka, M. J. Iliff, C. Lagoze, F. A. La Sorte, M. Merrifield, W. Morris, T. B. Phillips, M. Reynolds, A. D. Rodewald, K. V. Rosenberg, N. M. Trautmann, A. Wiggins, D. W. Winkler, W.-K. Wong, C. L. Wood, J. Yu, and S. Kelling. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. Biological Conservation 169:31-40.

Vander Haegen, W. M. 2007. Fragmentation by agriculture influences reproductive success of birds in a shrubsteppe landscape. Ecological Applications 17:934-947.

Vander Haegen, W. M., F. C. Dobler, and D. J. Pierce. 2000. Shrubsteppe bird response to habitat and landscape variables in eastern Washington, USA. Conservation Biology 14:1145-1160.

Vander Haegen, W. M., M. A. Schroeder, W-Y. Chang, and S. M. Knapp. 2015. Avian abundance and reproductive success in the intermountain west: Local-scale response to the conservation reserve program. Wildlife Society Bulletin 39:276-291.

WDFW. 2015. Washington's state wildlife action plan: 2015 update. Washington Department of Fish and Wildlife, Olympia.

Wiens, J. A., and J. T. Rotenberry. 1981. Habitat associations and community structure of birds in shrubsteppe environments. Ecological Monographs 51:21-41.

Wilson, K. A., M. I. Westphal, H. P. Possingham, and J. Elith. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation 122:99-112.

Wisdom, M. J., M. M. Rowland, and L. H. Suring, editors. 2005. Habitat threats in the sagebrush ecosystem: Methods of regional assessment and applications in the great basin. Alliance Communications Group, Lawrence, Kansas, USA.

APPENDIX. Tables and figures from models of species distribution for sagebrush obligate passerines in Washington.

Table A1. Parameter estimates from general linear model for predicting sagebrush sparrow occurrence in Washington. Variable names are defined in Table 2.

Variable	Estimate	SE	z	P
Intercept	-1.088	2.066	-0.527	0.5985
Sagebrush	0.121	0.032	3.789	0.0002
Litter	-0.131	0.052	-2.515	0.0119
Bare ground	-0.030	0.041	-0.744	0.4572
SSCRP_3km	-0.036	0.077	-0.459	0.6464
Longitude	-4.43E-06	1.18E-06	-3.752	0.0002
Bare ground:SSCRP_3km	3.16E-03	1.83E-03	1.731	0.0834

Table A2. General additive model for predicting sagebrush sparrow occurrence in Washington, with ANOVA sum of squares (SumSq) and Chi Square (ChiSq) statistics for parameter effects. Variable names are defined in Table 2.

Parametric effects				
Variable	Df	SumSq	F	P
Sagebrush	1	0.18	0.1123	0.7378
SSCRP_3km	1	20.17	12.2895	0.0005
s(Litter, 2)	1	6.11	3.7243	0.0546
s(Bare Ground, 8)	1	0.97	0.5928	0.4419
s(Longitude, 3)	1	4.3	2.6222	0.1065
Non-parametric effects				
Variable	Df	ChiSq	P	
s(Litter, 2)	1	3.8034	0.0511	
s(Bare Ground, 8)	7	19.569	0.0065	
s(Longitude, 3)	2	11.9602	0.0025	

Table A3. Parameter estimates for parametric effects in general additive model for predicting sagebrush sparrow occurrence in Washington. Variable names defined in Table 2.

Variable	Estimate	SE	z value	P
(Intercept)	-2.391	1.492	-1.602	0.1091
Sagebrush	0.079	0.034	2.305	0.0212
SSCRP_3km	0.102	0.028	3.615	0.0003

Table A4. Relative influence of explanatory variables from boosted regression tree model for predicting sagebrush sparrow occurrence in Washington. Variable names are defined in Table 2.

Variable	Relative variable influence
Bare Ground	32.767
SSCRP_3km	26.511
Latitude	16.314
Sagebrush	10.875
Litter	8.414
Soil Depth	3.575
Elevation	1.539
Agsoil	0.003

Table A5. Parameter estimates from general linear model for predicting sage thrasher occurrence in Washington. Variable names are defined in Table 2.

Variable	Estimate	SE	z value	P
Intercept	-6.802	0.883	-7.699	1.37E-14
Elevation	0.001	0.001	3.427	0.0006
Sagebrush	0.161	0.028	5.739	9.54E-09
Bare Ground	0.021	0.011	2.003	0.0452
SS_3km	0.095	0.026	3.584	0.0003
CRP_3km	0.175	0.050	3.487	0.0004

Table A6. General additive model for predicting sage thrasher occurrence in Washington, with ANOVA sum of squares (SumSq) and Chi Square (ChiSq) statistics for parameter effects. Variable names are defined in Table 2.

		Parametric effects		
Variable	Df	SumSq	F value	P
Sagebrush	1	23.001	22.330	3.58E-06
SS_3km	1	11.955	11.607	0.0008
CRP_3km	1	16.922	16.429	6.49E-05
s(Elevation, 4)	1	8.474	8.227	0.0044
Bare Ground	1	4.063	3.944	0.0479
		Non-parametric effects		
Parameter	Df	ChiSq	p	
s(Elevation, 4)	3	10.616	0.014	

Table A7. Parameter estimates for parametric effects in general additive model for predicting sage thrasher occurrence in Washington. Variable names are defined in Table 2.

Variable	Estimate	SE	z value	P
Intercept	-7.038	0.881	-7.982	1.43E-15
Sagebrush	0.173	0.029	5.944	2.78E-09
SS_3km	0.101	0.027	3.750	0.0002
CRP_3km	0.192	0.052	3.689	0.0002
Bare Ground	0.022	0.011	2.016	0.0438

Table A8. Relative influence of explanatory variables from boosted regression tree model for predicting sage thrasher occurrence in Washington. Variable names are defined in Table 2.

Variable	Relative importance value
Sagebrush	31.032
Elevation	23.918
SS_3km	17.461
Latitude	13.346
CRP_3km	6.3211
Litter	3.008
Soil Depth	2.649
Bare Ground	1.355
Longitude	0.871
Agsoil	0.036

Table A9. Parameter estimates from general linear model for predicting Brewer's sparrow occurrence in Washington. Variable names are defined in Table 2.

Variable	Estimate	SE	z value	P
Intercept	-0.771	1.26	-0.562	0.5741
Longitude	3.72E-06	1.08E-06	3.458	0.0005
Latitude	2.77E-06	8.48E-07	3.264	0.0011
Agsoil	-1.080	0.408	-2.658	0.0078
Elevation	0.002	3.04E-04	5.532	3.16E-08
Herbaceous	0.067	0.065	1.019	0.3081
Herbsq	-0.001	8.58E-04	-1.641	0.1007
Sagebrush	0.130	0.037	3.475	0.0005
Litter	-0.178	0.054	-3.3	0.0009

Table A10. General additive model for predicting Brewer's sparrow occurrence in Washington, with ANOVA sum of squares (SumSq) and Chi Square (ChiSq) statistics for parameter effects. Variable names are defined in Table 2.

Variable	Df	SumSq	F value	P
Longitude	1	2.349	2.399	0.1224
Agsoil	1	6.781	6.926	0.0089
Elevation	1	7.803	7.971	0.0051
Herbaceous	1	8.205	8.381	0.0041
Sagebrush	1	16.014	16.358	0.0007
Non-parametric effects				
Parameter	Df	ChiSq	p	
s(Latitude, 3)	2	7.220	0.027	
s(Litter, 2)	1	4.543	0.033	

Table A11. Parameter estimates for parametric effects in general additive model for predicting Brewer's sparrow occurrence in Washington. Variable names are defined in Table 2.

Variable	Estimate	SE	z value	P
(Intercept)	-0.152	1.132	-0.1341	0.8933
Longitude	2.667E-06	1.064E-06	2.5057	0.0122
Agsoil	-0.988	0.414	-2.384	0.0171
Elevation	0.002	0.0003	5.2084	<0.0001
Herbaceous	-0.033	0.015	-2.2386	0.0252
Sagebrush	0.137	0.037	3.7676	0.0002

Table A12. Relative influence of explanatory variables from boosted regression tree model for predicting Brewer's sparrow occurrence in Washington. Variable names are defined in Table 2.

Variable	Relative Importance Value
Latitude	59.111
Elevation	15.913
Sagebrush	14.173
Soil depth	4.5180
Litter	1.886
SS_3km	1.484
CRP_3km	1.109
Bare Ground	0.926
Herbaceous	0.565
Longitude	0.278
Agsoil	0.034
Herbsq	0

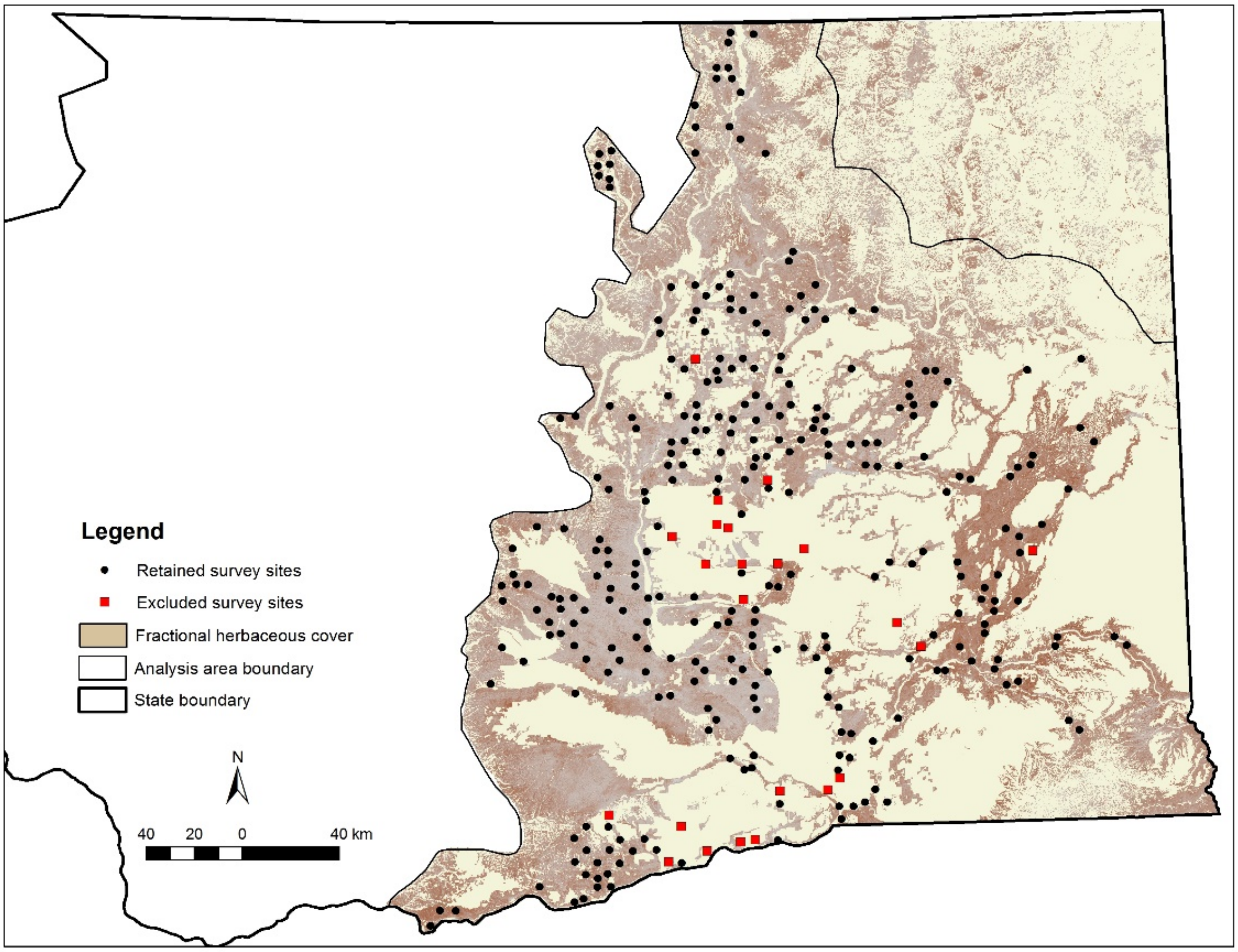


Figure A1. Surveyed sites retained for modeling of shrubsteppe bird distribution in Washington or excluded due to lack of overlap with modeled vegetation layers, 2014-2019.

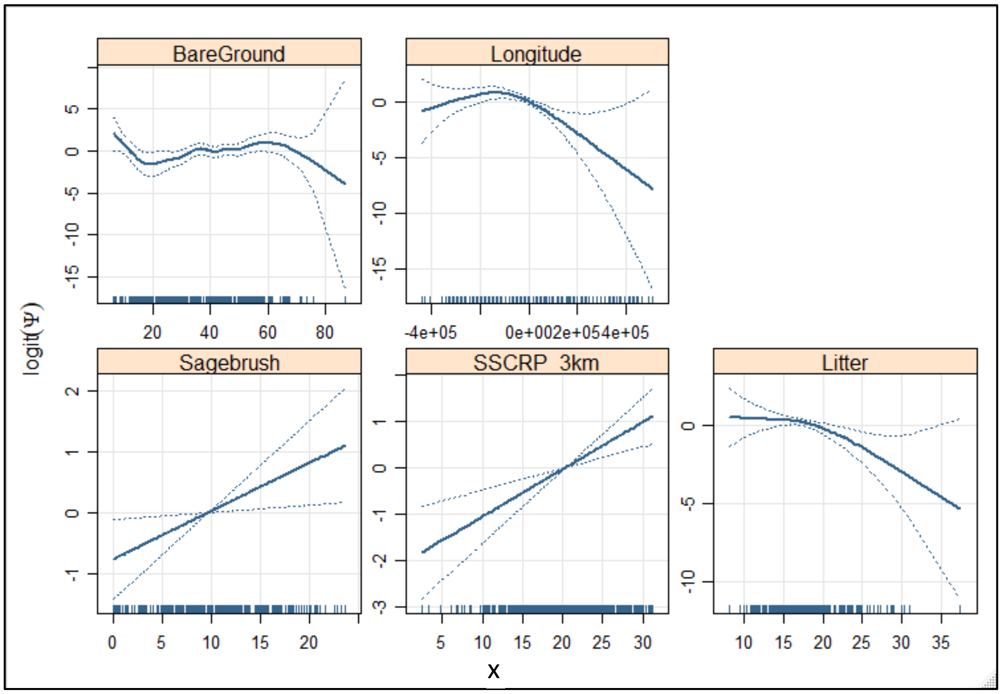


Figure A2. Linear and non-linear effects from generalized additive model for sagebrush sparrow distribution in eastern Washington, 2014-2019.

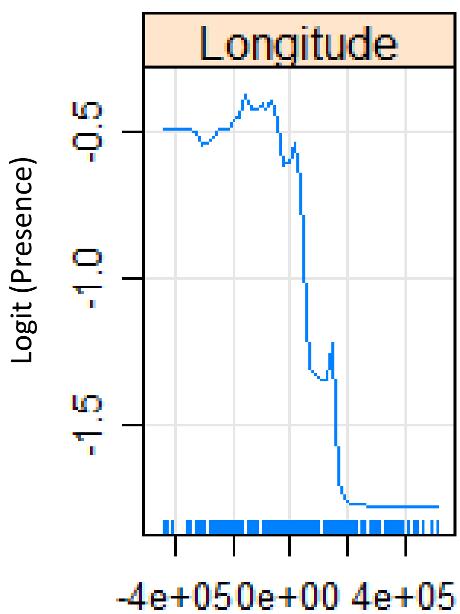


Figure. A3. Partial dependence plot for variable Longitude from random forest model of sagebrush sparrow distribution in Washington, 2014-2019.

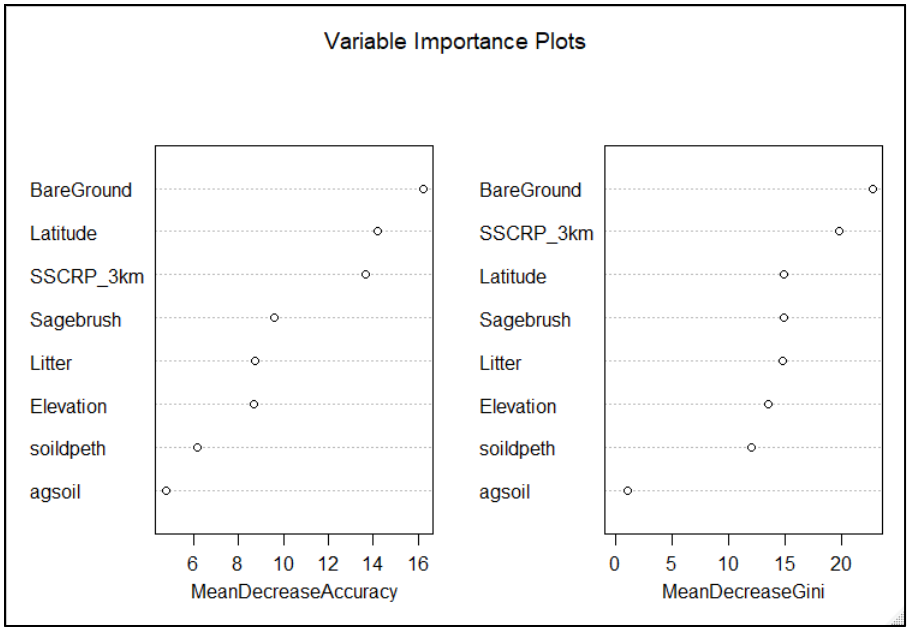


Figure A4. Variable importance plots from random forest model for sagebrush sparrows in study of shrubsteppe bird distribution in Washington, 2014-2019.

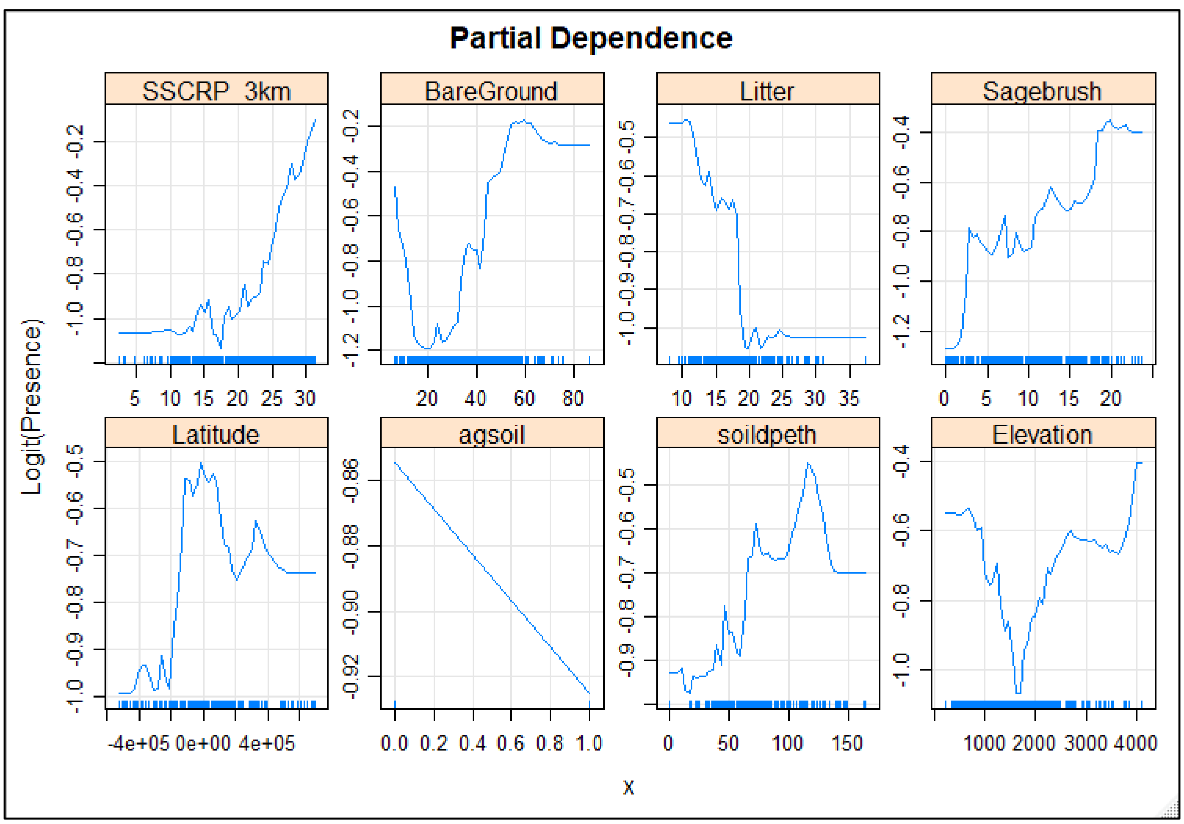


Figure A5. Partial dependence plots from random forest model for sagebrush sparrow distribution in eastern Washington, 2014-2019.

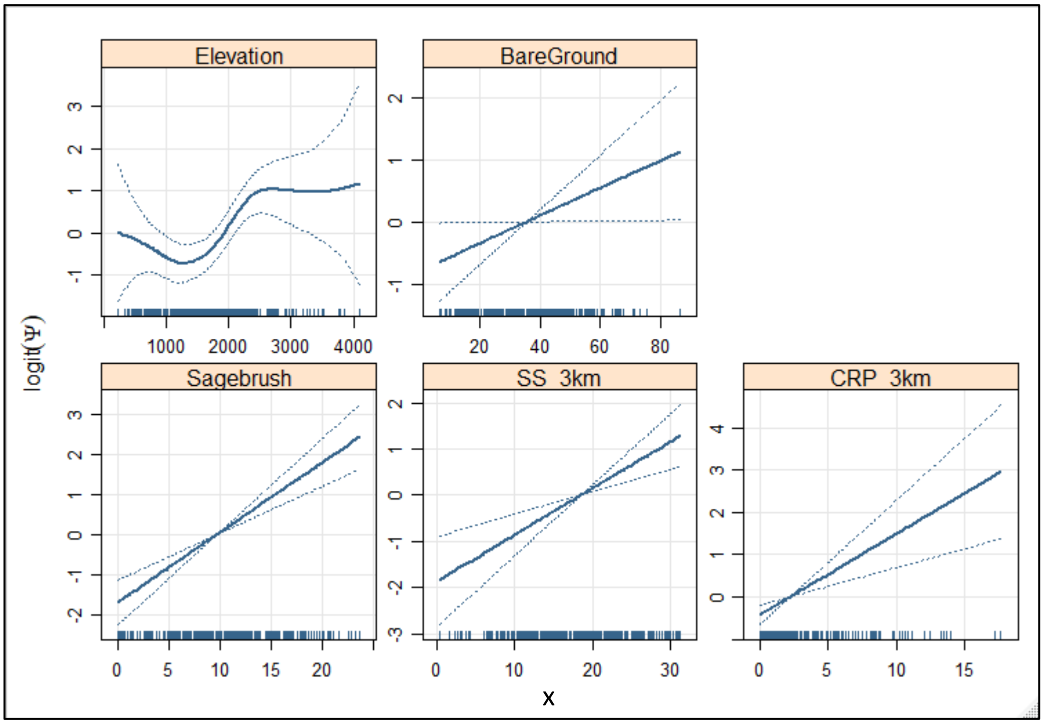


Figure A6. Linear and non-linear effects from generalized additive model for sage thrasher distribution in eastern Washington, 2014-2019.

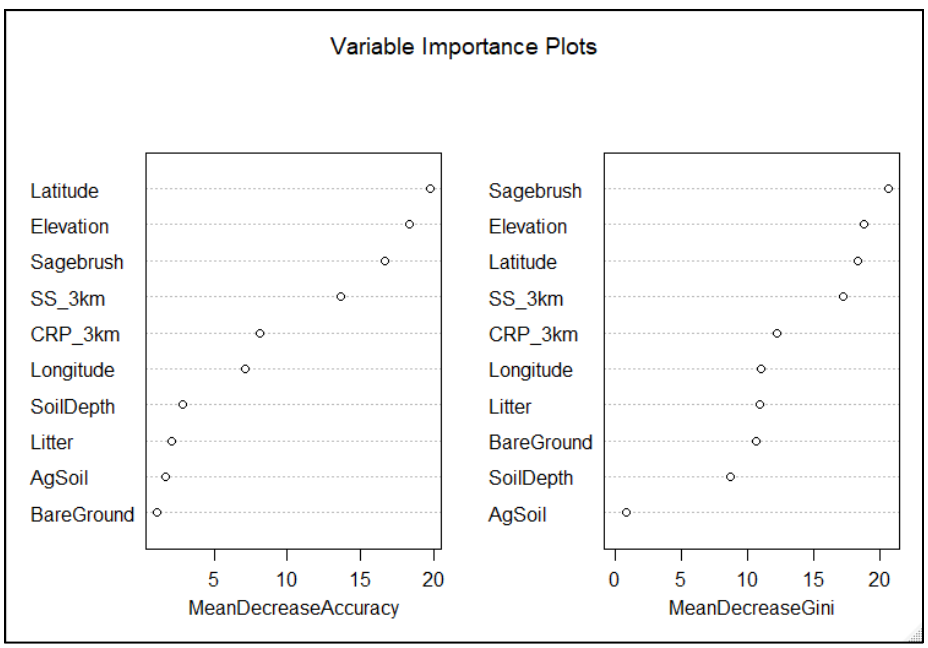


Figure A7. Variable importance plots from random forest model for sage thrashers in study of shrubsteppe bird distribution in Washington, 2014-2019.

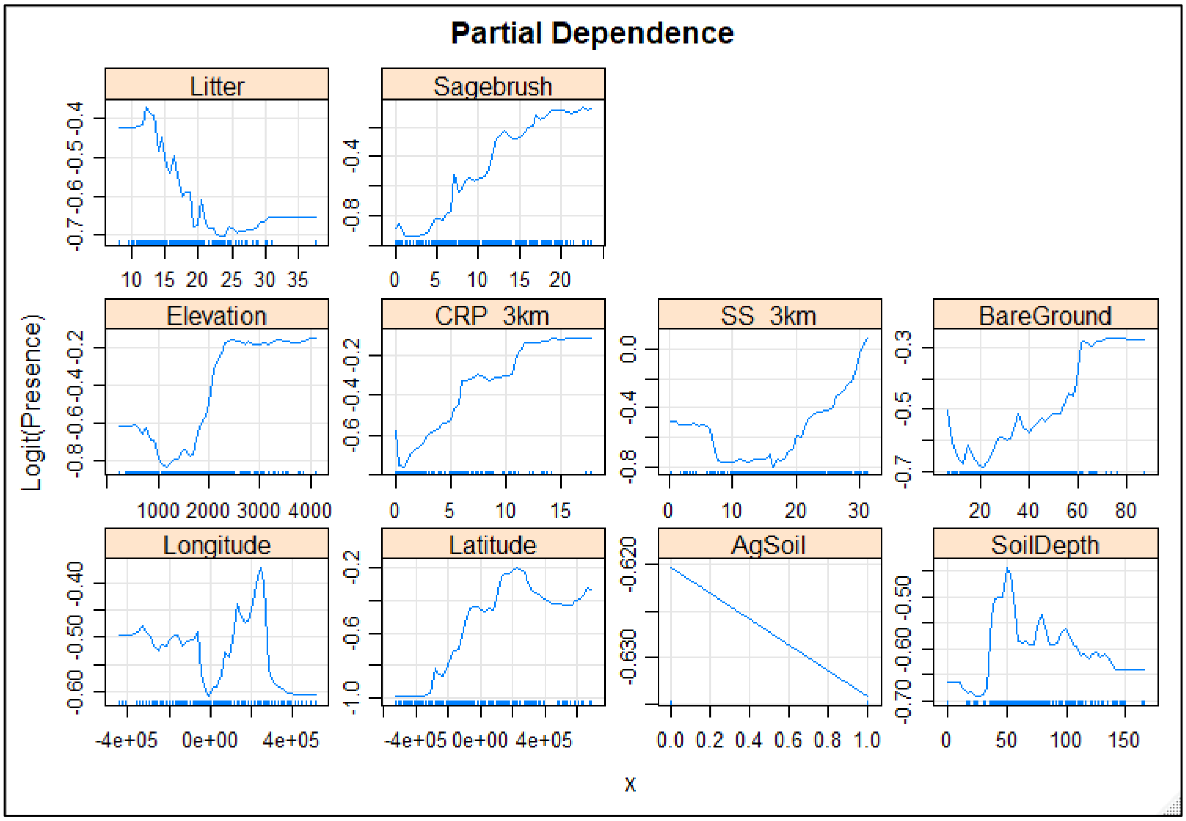


Figure A8. Partial dependence plots from random forest model for sage thrasher distribution in eastern Washington, 2014-2019.

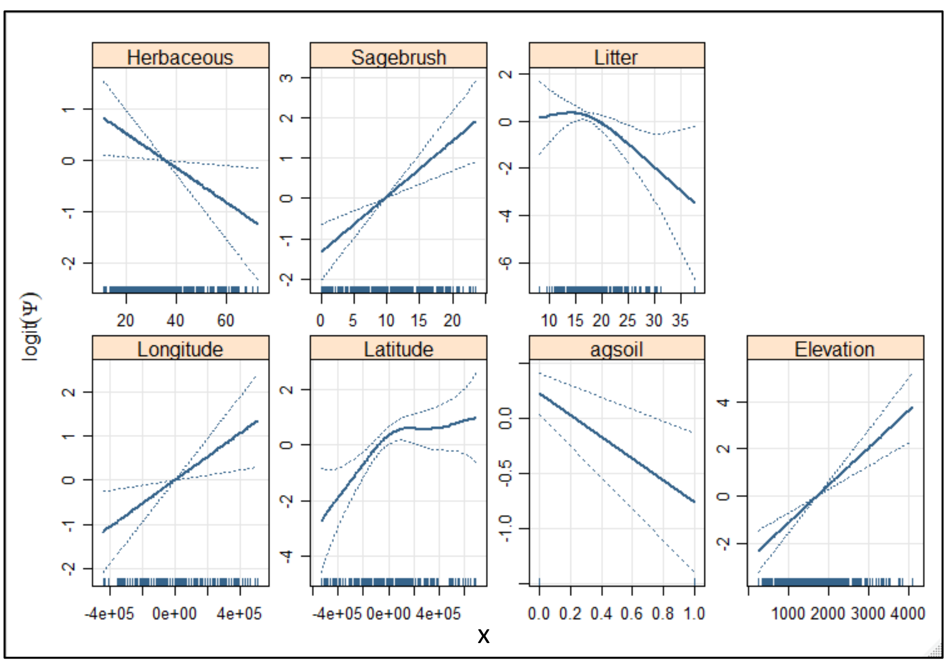


Figure A9. Linear and non-linear effects from generalized additive model for Brewer's sparrow distribution in eastern Washington, 2014-2019.

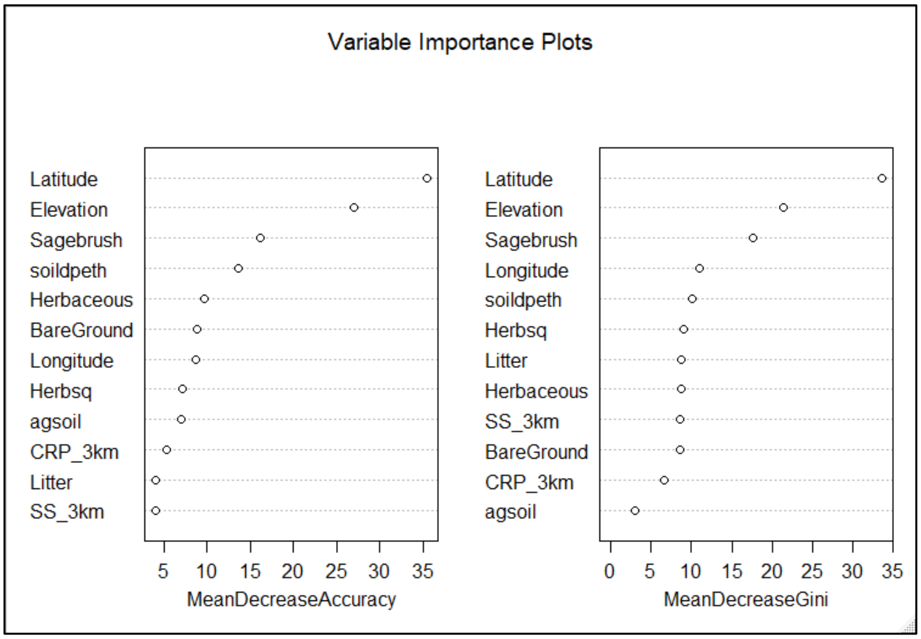


Figure A10. Variable importance plots from random forest model for Brewer's sparrows in study of shrubsteppe bird distribution in Washington, 2014-2019.

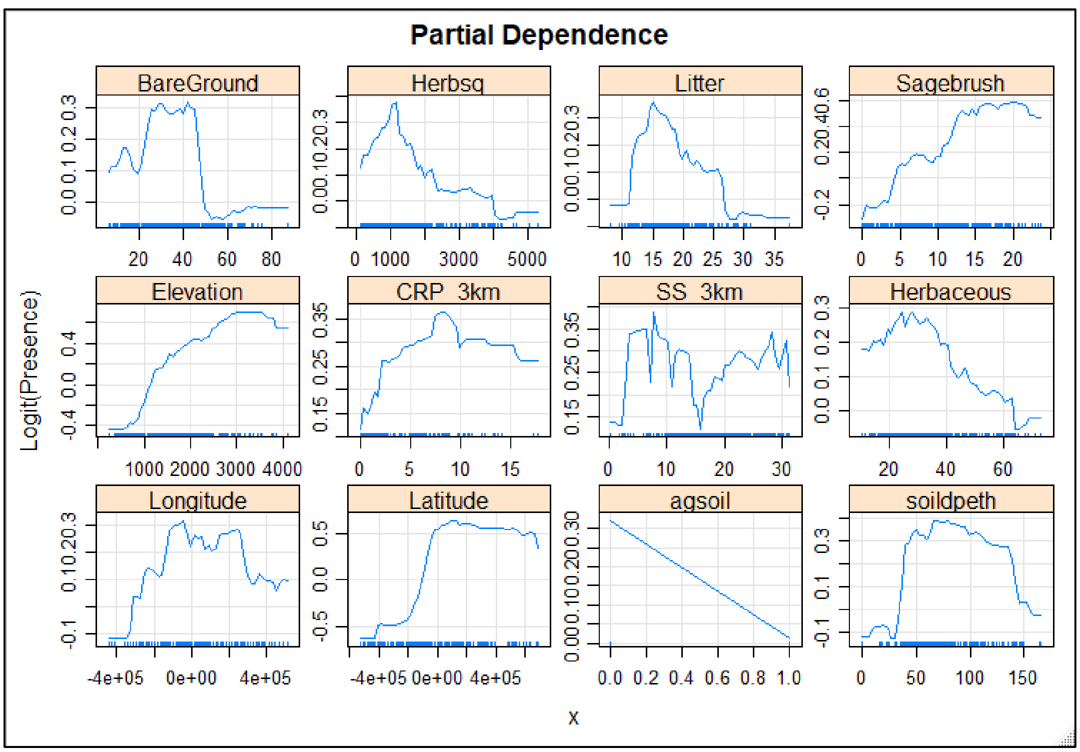


Figure A11. Partial dependence plots from random forest model for Brewer's sparrow distribution in eastern Washington, 2014-2019.