A Comparison of Remotely Sensed Environmental Predictors for Avian Distributions

Laurel M. Hopkins[1], Tyler A. Hallman[2,4], John Kilbride[3], W. Douglas Robinson[4], and Rebecca A. Hutchinson[1,4]

[1]Department of Electrical Engineering and Computer Science, Oregon State University, 1148 Kelley Engineering Center, Corvallis, OR 97331, USA

[2]Monitoring Department, Swiss Ornithological Institute, Seerose 1, Sempach, Switzerland CH-6204

[3]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 331 Strand Hall, Corvallis, OR, 97331, USA

[4]Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, 2820 Nash Hall, Corvallis, OR, USA

Corresponding Author:
Laurel Hopkins
Tel: +1 541-737-3617
Email: hopkilau@oregonstate.edu

Author ORCIDs:
Laurel Hopkins: 0000-0002-7689-4827
Tyler A. Hallman: 0000-0003-2604-0548
John Kilbride: 0000-0002-9421-6480
W. Douglas Robinson: 0000-0003-2240-0606

Abstract - 250 words

Context - 36

With greater accessibility and processing power from online platforms, summaries of remotely sensed data are increasingly used in species distribution models (SDMs). Comparisons of the predictive power of these environmental variables could inform SDMs moving forward.

Objectives - 85

We evaluated the performance of freely available Landsat data as predictor sets for SDMs. Our objectives were to 1) compare the performance of single season SDMs built on mean values of raw spectral bands, Tasseled Cap transformations, and eight different indices, including NDVI, 2) evaluate the performance gain with the addition of standard deviation, textural metrics, and additional seasons, and 3) compare the performance of SDMs built on these continuous spectral predictor sets to SDMs built on classified land cover data (e.g., percent forest cover).

Methods - 31

We used statewide point counts to build multi-scale SDMs for 13 avian species across Oregon, USA. We compared the performance of SDMs built on each predictor set based on our objectives.

Results - 60

Of the Landsat-derived predictor sets, SDMs built on raw spectral bands had the highest overall performance with nearly equivalent performance in Tasseled-Cap models. While performance gains from standard deviations, textural metrics, and additional seasons were minimal in raw-band and Tasseled-Cap models, gains were appreciable in single-index models. Classified land cover models performed equivalently to raw band models.

Conclusions - 36

When predictive performance is paramount, means of raw Landsat bands are strong predictors for avian SDMs. When parsimonious variables are essential, SDMs of single indices (e.g., NDVI) greatly benefit from additional information, such as standard deviation.

Keywords

Remote Sensing, Species Distribution Models, Landsat, Google Earth Engine

Acknowledgements

Laurel Hopkins received funding from the National Aeronautics and Space Administration under Future Investigators in NASA Earth and Space Science and Technology (FINESST) (Grant No: 80NSSC20K1664). This work was also supported in part by the National Science Foundation under project NSF IIS 1910118.

Introduction

Remotely sensed data have been essential to characterizing the environment for species distribution models (SDMs) for decades (Kerr et al. 2001; Gottschalk et al. 2005; Cord et al. 2013; Randin et al. 2020). Increased accessibility and processing power through free platforms like Google Earth Engine (Gorelick et al. 2017) have catalyzed a proliferation in the use of summaries of satellite imagery. A myriad of remotely sensed datasets and summary methods have been used in SDMs (e.g., Kerr et al. 2001; Buermann et al. 2008; Shirley et al. 2013) which has led to an immense number of satellite-derived environmental predictor variables available to researchers. As few comparative studies exist (e.g., Cord et al. 2014), there is a need to compare and document the relative value of each to SDMs. For the purposes of comparison, we categorize remotely sensed data into two main groups, both of which are commonly used to inform SDMs (Kerr and Ostrovsky 2003, He et al. 2015): raw satellite images, herein "unclassified data", which retain the continuous nature of the images (e.g., means of the image bands, texture metrics) and classified data, which are derived from satellite images and map image pixels into discrete categories (e.g., land cover classes). This paper compares the predictive performance of unclassified Landsat-derived environmental predictor variables in SDMs and also evaluates how these predictor variables compare to those developed from classified land cover datasets.

Landsat, a multispectral satellite dataset commonly used in SDMs (Kerr and Ostrovsky 2003; Gottschalk et al. 2005), has acquired Earth observations continuously since 1972. It has an update cycle of 16-days and a 30 m spatial resolution. While categorized as moderate resolution multispectral imagery, for use in predicting distributions of most wildlife, this is considered a high-resolution dataset. The temporal and spatial resolution of Landsat data and the extensive historical archive of radiometrically and geometrically calibrated imagery make Landsat data appealing for modeling ecosystem processes (Kennedy et al. 2014; Wulder et al. 2019). Indeed, Landsat observations have been used extensively to characterize ecosystem structure and processes (e.g., Foody et al. 1996, Pflugmacher et al. 2012, Baumann et al. 2017, Meigs et al. 2020).

Raw spectral bands from Landsat can inform SDMs through either direct summarization (Gottschalk et al. 2005; Shirley et al. 2013) or the computation of indices and transformations (Osborne et al. 2001; Seto et al. 2004; Buermann et al. 2008; Parviainen et al. 2013). When working with modeling methods that benefit from fewer predictor variables, it may be advantageous to use indices, which are single values computed from the raw bands that characterize physical attributes of the landscape. For example, the normalized difference vegetation index (NDVI), which describes the spectral relationship between red reflectance and near-infrared reflectance and is a proxy for photosynthetic activity, is commonly used in SDMs (Krishnaswamy et al. 2009; Bradley and Fleishman 2008; Osborne et al. 2001; Seto et al. 2004). Though not as frequently as NDVI, other indices such as the normalized difference snow index (NDSI) and enhanced vegetation index (EVI) have also been used in SDMs (Cord et al. 2014; Niittynen et al. 2018). The Tasseled Cap transformation is a common dimensionality reduction technique for spectral imagery that has also been used to inform SDMs (Zimmermann et al. 2007; Oeser et al 2020). The Tasseled Cap transformation is a reprojection of the raw bands into three dimensions representing brightness, greenness, and wetness (Crist and Cicone 1984). It is also possible to characterize temporal variation by summarizing remotely sensed imagery over seasons or to describe spatial variation in vegetation with textural metrics. Deriving the reflectance values for multiple seasons (e.g., spring, summer, fall) may allow SDMs to further differentiate available habitats compared to SDM inputs derived from a single season (i.e., an image obtained during the peak of vegetation phenology or during a pre-defined breeding season). Recently, texture metrics have been shown to be strong predictors of bird distributions (Farwell et al. 2020).

An alternative to unclassified data are classified land cover datasets. Classified land cover datasets are produced by mapping the raw spectral values at each pixel into discrete land cover classes (e.g., urban, grassland, deciduous forest). For example, the National Land Cover Dataset (NLCD) (Dewitz 2019), which is derived from Landsat, consists of 20 classes which span categories such as human-developed, forests, and wetlands. A limitation to classified datasets is that they are generally released annually or every few years whereas unclassified datasets are updated at intervals measured in days. The finer temporal resolution of the unclassified datasets allows for faster detection of environmental changes. Further, environmental information is lost due to the coarse aggregation of continuous spectral bands into discrete classes (Foody 2002; Gottschalk et al. 2005; Gillespie et al. 2008; Krishnaswamy et al. 2009), which may limit the predictive performance of SDMs (Bradley and Fleishman 2008). Conversely, compositional information (i.e., proportions of land cover classes) may be gained with summaries of classified data. While summaries of raw spectral values may imply which types of land cover are more prevalent (e.g., a very green landscape is more likely covered in trees rather than water or barren land), they do not explicitly identify land cover classes. For species that prefer specific habitat types, knowing what proportion of an environment is composed of specific habitats may be more informative than summaries of raw spectral values. Due to ease of interpretation, ecologists frequently use classified data to inform SDMs. For example, Johnston et al. (2021) suggests the use of classified land cover for developing environmental variables when modeling citizen science species records, such as eBird.

This paper compares the performance of SDMs trained on sets of habitat variables derived from Landsat imagery for several bird species in the state of Oregon, USA. Our primary objectives were to 1) identify the indices or transformations of raw bands that consistently informed the highest performing SDMs across species, 2) examine whether data from additional seasons improved SDM performance, and 3) explore whether standard deviations or textural metrics improved SDM performance. Our secondary objective was to compare the performance of SDMs built on unclassified Landsat data to the performance of SDMs built on commonly used classified landcover datasets (NLCD and MODIS). Finally, we offer suggestions for applying remotely sensed data to SDMs with the goal of helping guide researchers through the many options faced when selecting remotely sensed data for SDMs.

Methods

Study area

We used the state of Oregon, in the Pacific Northwest of the United States of America as our study area. Oregon's 255,026 km^2 area includes nine distinct ecoregions, 12 Köppen climate types, an elevational range from sea level to more than 3500 m, and habitats from densely populated cities to remote wilderness.

Bird surveys and sampling design

We used bird surveys conducted through the Oregon 2020 project (Robinson et al. 2020). Count locations in Oregon 2020 were distributed across Oregon in a stratified random manner (Robinson et al. 2020). The strata were defined by the Public Land Survey System, which divides the state into 6 × 6-mile townships, generating a total of 36 square-mile sections within each of the more than 2800 townships. Robinson et al. (2020) selected at random a one square-mile section from each township. They kept that section if it had some form of public access such as a road or trail. If there was no access, they shifted the section to the next nearest section that had access and similar habitat type and elevation as determined from inspection of Google Earth imagery. Within each section, they conducted point counts approximately every 200 m along publicly accessible roads or trails. The modal number of locations sampled within each square was four and ranged from one to 12. Robinson et al. (2020) supplemented this statewide sampling design with additional surveys conducted at 0.8-km intervals along nearly every accessible road in Benton and Polk Counties. They also included surveys conducted in a 200-m grid established across the William L. Finley National Wildlife Refuge in Benton County. Robinson et al. (2020) showed that the proportional coverage of habitats available in Oregon were extremely similar to those covered by their point sampling scheme.

Three trained observers conducted 10,844 5-minute stationary surveys during the breeding seasons (May 15th to July 10th), 2011-2019 (Robinson et al. 2020). All surveys were performed between dawn and noon, unless bird activity noticeably declined earlier. At each survey, all birds detected were identified to species. All sites were visited once. Distance sampling and time of detection methods were implemented in counts to allow for direct estimation of imperfect detection, but to simplify analyses and findings, and to better mirror commonly available eBird data, which contain no such ancillary methods, we removed these data and did not account for imperfect detection in our models. As our interest primarily lies in the comparative performance of environmental predictor sets, and imperfect detection derived from standard variables such as time of day and day of year should bias all models equivalently (e.g., all built on the same data), comparisons should remain unaffected. Additionally, we reduced species counts to detections and non-detections. Robinson et al. (2020) provide further details.

Species selection

To represent a range of habitat types and levels of species' habitat specializations, we selected thirteen species occurring in six common habitat types in Oregon (Table 1). All of these species were detected effectively by the sampling method in Robinson et al. (2020) since they vocalize frequently during the breeding season. For each habitat, we included two or three species; One species was considered to be a generalist and one or two were considered specialists based on our own experiences and qualitative data (Marshall et al. 2003). Generalists often occupy a primary habitat and also other structurally similar habitats, so we anticipated relationships between remotely sensed habitat data and species occurrence would be weaker than for specialist species and their habitats.

Table 1 Study species, the primary habitat(s) they occupy, whether we considered them to be generalists or specialists on the primary habitat type, and their sample prevalence in our Oregon 2020 data. Our usage of generalist and specialist are relative to the species in the study.

Species	Primary habitat	Specialist or Generalist	Prevalence
Western Tanager ***Piranga ludoviciana***	Forest	Generalist	0.2478
Hermit Warbler ***Setophaga occidentalis***	Coniferous forest canopy	Specialist	0.2020
Pacific Wren ***Troglodytes pacificus***	Coniferous forest understory	Specialist	0.1350
Sage Thrasher ***Oreoscoptes montanus***	Sagebrush	Generalist	0.0432
Sagebrush Sparrow ***Artemesiospiza nevadensis***	Mature sagebrush	Specialist	0.0247
Swainson's Thrush ***Catharus ustulatus***	Moist woodlands	Generalist	0.2850
Hermit Thrush ***Catharus guttatus***	Higher elevation woods	Specialist	0.0601
Western Meadowlark ***Sturnella neglecta***	Grassland/sagebrus h	Generalist	0.1621
Savannah Sparrow ***Passerculus sandwichensis***	Grassland	Specialist	0.0775
Yellow Warbler ***Setophaga petechia***	Riparian woods	Generalist	0.0538
Yellow-breasted Chat ***Icteria virens***	Riparian/shrubs	Specialist	0.0350
Ash-throated Flycatcher ***Myiarchus cinerascens***	Juniper/oaks	Generalist	0.0213
Gray Flycatcher ***Empidonax wrightii***	Juniper	Specialist	0.0468

Remotely sensed data and spectral predictor sets

The basis for our analysis was three time-series of gap-free, radiometrically-consistent composited satellite imagery from which we computed spectral predictor sets. An overview of the image processing workflow is shown in Figure 1. First, we developed three time-series of composited imagery, one each for the spring, summer, and fall seasons, using Landsat satellite imagery. Then, using the LandTrendr algorithm, we processed the annual composites into a time-series of gap-free, radiometrically consistent images (Kennedy et al. 2010). Using these stabilized time-series, we computed ten spectral datasets: raw bands, Tasseled Cap transformations, and eight single indices across the study region. Finally, we calculated summaries (e.g., means) of the spectral datasets over buffers with multiple radii centered at the bird count locations for all three seasons. See Online Resource 1 for a more detailed description of the image processing workflow.

The spectral datasets we selected build off of those from past species distribution models (Gottschalk et al. 2005; Buermann et al. 2008; Shirley et al. 2013; Oeser et al 2020). Specifically, we summarized raw spectral bands, their associated Tasseled Cap transformations, and eight single-valued indices derived from the raw bands: NDMI, NDVI, NBR, NBR2, EVI, SAVI, MSAVI, NDSI (Table 2). The Tasseled Cap transformation is computed by projecting the spectral bands into three dimensions, or spectral indicators, that describe brightness, greenness, and wetness (Crist and Cicone 1984). We also included the Tasseled Cap Angle (TCA), as a fourth variable in the Tasseled Cap predictor set (Table 2). We selected the single-valued indices as they are frequently used in ecological remote sensing and are readily available to researchers as part of the Landsat Collection 1 Surface Reflectance data produced from the USGS. In this analysis, we specifically examined single-date remote sensing metrics (i.e., they were computed using a single image) to constrain the number of predictors being investigated (Seto et al. 2004; Meddens et al. 2013).

For every count location and each of the ten spectral datasets, we constructed spectral predictor sets by calculating summaries over the buffered regions for all three seasons. Specifically, we calculated the mean and standard deviation of each of the bands in the spectral datasets with 75, 600, and 2400 m radii buffers centered at the count location for spring, summer, and fall imagery (Table 3). Species respond to their environments at different scales (Wiens and Milne 1989). The use of multiple buffers to characterize environmental covariates can ensure that a species-specific appropriate environmental scale is included (Hallman and Robinson 2020a). We selected a range of buffers that have previously been shown to predict songbirds (Hallman and Robinson 2020a; Hallman and Robinson 2020b; Hallman et al. 2021). We matched the year of the species observation to the year in which the Landsat imagery was collected. In addition to the means and standard deviations of each of the bands, we calculated seven GLCM texture metrics (Table 4) for all three seasons at all three buffer radii. Like standard deviations of bands, GLCM texture metrics characterize textural information (i.e., spatial arrangement) and have been shown to be informative of bird richness (Farwell et al. 2020).

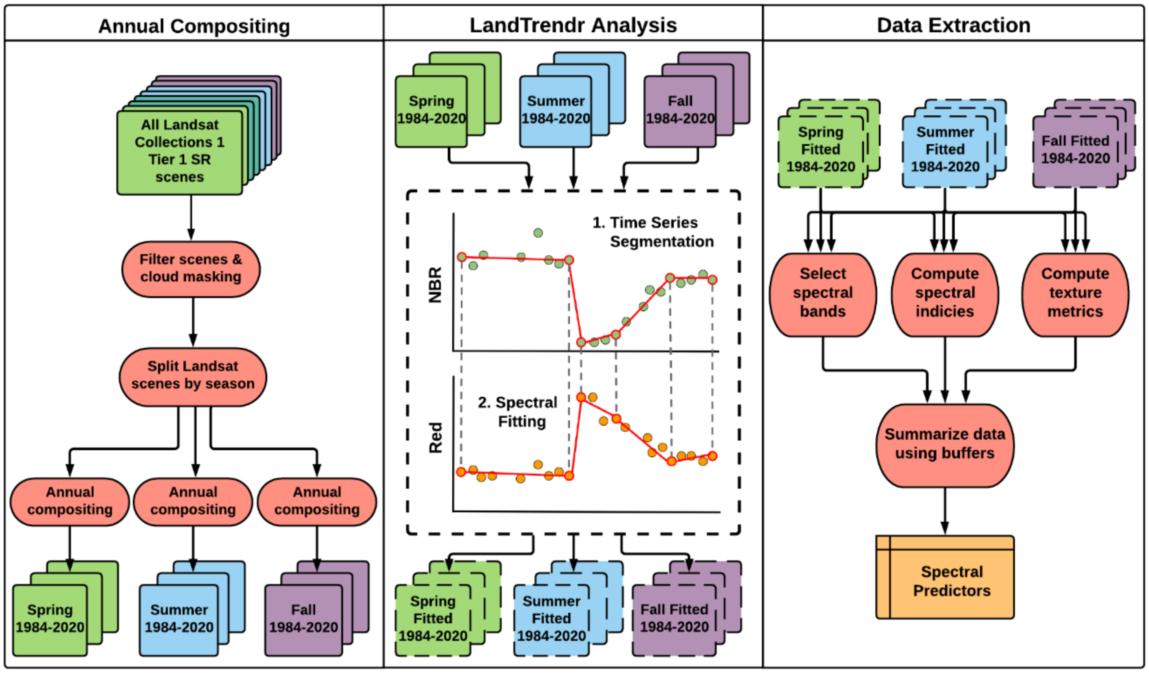

Fig. 1 *Flowchart of Landsat image processing using LandTrendr algorithm*

Table 2 A description of the spectral bands and indices that were used in the analysis. Each of the 18 metrics were computed for each of the fitted seasonal satellite image time series. The Landsat TM/ETM+ band naming conventions are used when describing how each metric was calculated.

Name (abv.)	**Description**	**Calculation**	**Source**
Visible Blue (B1)	Blue reflectance. Landsat TM/ETM+ band 1 (0.45-0.52 μm); Landsat OLI band 2 (0.45 - 0.51 μm)	-	-
Visible Green (B2)	Green reflectance. Landsat TM/ETM+ band 2 (0.52-0.60 μm); Landsat OLI band 3 (0.53 - 0.59 μm)	-	-
Visible Red (B3)	Red reflectance. Landsat TM/ETM+ band 3 (0.63-0.69); Landsat OLI band 4 (0.64 - 0.67 μm)	-	-
Near-infrared (B4)	Near-infrared reflectance. Landsat TM/ETM+ band 4 (0.76 - 0.90 μm); Landsat OLI band 5 (0.85 - 0.88 μm)	-	-
Short Wavelength Infrared 1 (B5)	Shortwave-infrared reflectance. Landsat TM/ETM+ band 5 (1.55 - 1.75 μm); Landsat OLI band 6 (1.57 - 1.65 μm)	-	-

Short Wavelength Infrared 2 (B7)	Shortwave-infrared reflectance 2. Landsat TM/ETM+ band 7 (2.08 - 2.35 μm); Landsat OLI band 7 (2.11 - 2.29 μm)	-	-
Tasseled Cap Brightness (TCB)	TCB captures the total reflectance exhibited by a location. Changes in soil condition produce large changes in TCB.	0.2043 * B1 + 0.4158 * B2 + 0.5524 * B3 + 0.5741 * B4 + 0.3124 * B5 + 0.2303 * B7	Crist and Cicone 1984
Tasseled Cap Greenness (TCG)	TCG is sensitive to changes in red and near-infrared reflectance associated with green vegetation.	-0.1603 * B1 + -0.2819 *B2 + -0.4934 * B3 + 0.794 * B4 + -0.0002 * B5 + -0.1446 * B7	Crist and Cicone 1984
Tasseled Cap Wetness (TCW)	TCW is responsive changes in soil and canopy moisture content, particularly changes that are expressed the shortwave infrared bands.	0.0315 * B1 + 0.2021 * B2 + 0.3102 * B3 + 0.1594 * B4 + -0.6806 * B5 + -0.6109 * B7	Crist and Cicone 1984
Tasseled Cap Angle (TCA)	Characterizes the proportion of vegetated to non-vegetated area within a pixel (White et al. 2011).	Arctan(TCG / TCB)	Powell et al. 2010
Normalized Difference Vegetation Index (NDVI)	NDVI is used as a proxy for vigor. The spectral index exploits the "red-edge" effect exhibited by green vegetation caused by the absorption of photosynthetically active radiation.	(B4 – B3) / (B4 + B3)	Rouse et al. 1974
Normalized Difference Moisture Index (NDMI)	The NDMI captures changes in moisture conditions on the ground and in the vegetation canopy.	(B4 – B5) / (B4 + B5)	Hardisky et al. 1983; Wilson and Sader 2002
Normalized Burn Ratio (NBR)	High NBR values indicate a strong soil signal and a lack of vegetation. Greater biomass densities decrease the soil signal and produces lower NBR values.	(B4 - B7) / (B4 + B7)	Key and Benson 1999
Normalized Burn Ratio 2 (NBR2)	NBR2 is a modification of the NBR which replaces B4 with B5. NBR2 is designed to capture variations in canopy moisture content during post-fire recovery.	(B5 - B7) / (B5 + B7)	Key and Benson 2006
Enhanced Vegetation index (EVI)	An optimization of the NDVI which attempts to decouple the background canopy signal from the soil signal and to account for changes in atmospheric conditions.	2.5 * ((B4 – B3) / (B4 + 6 * B3 – 7.5 * B1 + 1))	Huete et al. 1999
Soil Adjusted Vegetation Index (SAVI)	The SAVI corrects the NDVI for the influence of soil brightness in low vegetation cover areas.	((B4 – B3) / (B4 + B3 + 0.5)) * (1.5)	Huete 1988
Modified Soil Adjusted Vegetation Index (MSAVI)	The MSAVI is an optimization of the SAVI designed to further reduce influence of the background soil signal.	(2 * B4 + 1 – sqrt ((2 * B3 + 1)^2 – 8 * (B4 – B3))) / 2	Qi et al. 1994
Normalized Difference Snow Index (NDSI)	The NDSI exploits the difference between green and shortwave infrared reflectance exhibited by snow and ice.	(B2 – B5) / (B3 + B5)	Hall et al. 1995

Table 3. Example of spectral predictor sets for mean summer values. As an additional example, the raw bands Sp/Su/Fa means and standard deviations spectral predictor set contains 108 variables (18 means and 18 standard deviations for each of the three seasons).

Spectral dataset	Image bands	Buffer radii	Season	Summary method	Total # of variables in spectral predictor set
Raw bands	B1, B2, B3, B4, B5, B7	75, 600, 2400 m	Summer	Mean	18
Tasseled Cap	TCB, TCG, TCW, TCA	75, 600, 2400 m	Summer	Mean	12
Normalized Difference Vegetation Index (NDVI)	NDVI	75, 600, 2400 m	Summer	Mean	3
Normalized Difference Moisture Index (NDMI)	NDMI	75, 600, 2400 m	Summer	Mean	3
Normalized Burn Ratio (NBR)	NBR	75, 600, 2400 m	Summer	Mean	3
Normalized Burn Ratio 2 (NBR2)	NBR2	75, 600, 2400 m	Summer	Mean	3
Enhanced Vegetation index (EVI)	EVI	75, 600, 2400 m	Summer	Mean	3
Soil Adjusted Vegetation Index (SAVI)	SAVI	75, 600, 2400 m	Summer	Mean	3
Modified Soil Adjusted Vegetation Index (MSAVI)	MSAVI	75, 600, 2400 m	Summer	Mean	3
Normalized Difference Snow Index (NDSI)	NDSI	75, 600, 2400 m	Summer	Mean	3

Table 4 A description of the textural metrics calculated for this analysis. The notation is adopted from Haralick and Shanmugam (1974): *p(i,j)* is the gray-tone spatial dependence matrix calculated for a given angular offset. The terms μ_x, μ_y, σ_x, and σ_y describe the mean and standard deviation of the marginal probability distributions $P_x(i)$ and $P_y(j)$ (see Welch et al. (1988) for details).

Name (abv.)	Description	Calculation	Source
Contrast	A measure of the average amount of local variation (Haralick and Shanmugam 1974).	$= \sum_{n=0}^{N_g-1} n^2 \{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\}$	Haralick et al. 1973

Correlation	Characterizes linear gray-tone dependencies (Haralick and Shanmugam 1974).	$= \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$	Haralick et al. 1973
Variance	Measures the dispersion of the of values in the GLCM matrix (Welch et al. 1988).	$= \sum_i \sum_j (i - \mu)^2 p(i,j)$	Haralick et al. 1973
Entropy	Describes the randomness of values in the image (Welch et al. 1988).	$= -\sum_i \sum_j p(i,j) \log [p(i,j)]$	Haralick et al. 1973
Inertia	Measures the spread of values in the GLCM matrix (Welch et al. 1988).	$= \sum_i \sum_j (i - j)^2 p(i,j)$	Conners et al. 1984
Shade	Quantifies the skewness of the distribution of values in the GLCM matrix (Welch et al. 1988).	$= \sum_i \sum_j \left(i + j - \mu_x - \mu_y\right)^3 p(i,j)$	Conners et al. 1984
Prominence	Quantifies the tailedness of the GLCM matrix (Welch et al. 1988).	$= \sum_i \sum_j \left(i + j - \mu_x - \mu_y\right)^4 p(i,j)$	Conners et al. 1984

In addition to the unclassified imagery, we summarized two classified datasets to evaluate how the unclassified spectral predictor sets compare to those developed from classified imagery. Johnston et al. (2021) recommend creating environmental variables by summarizing MCD12Q1 v006 (Friedl and Sulla-Menashe 2015), a classified MODIS dataset, by calculating the proportion of each class present in 2.5 × 2.5 km kernels centered at species records. Because this type of use of habitat composition data is so common, we focus on the compositional aspects of remote sensed data but acknowledge that configuration variables also may contribute to accurate prediction of species distributions (Mazerolle and Villard 1999). The spatial resolution of MCD12Q1 is 500 × 500 m which is much larger than the 30 × 30 m resolution of Landsat imagery and the resulting spectral predictors. To maintain the same spatial resolution across unclassified and classified data, we computed summaries from NLCD2016, a classified dataset derived from Landsat which has the same 30 × 30 m resolution. A limitation to NLCD is that it only contains data for the United States compared to MCD12Q1's global coverage. We calculated the proportion of land cover classes present for the three buffer radii, as is commonly done when summarizing classified data (Thuiller et al. 2004; Johnston et al. 2021). We also computed summaries of the much coarser resolution MCD12Q1 to compare to the best practices of Johnston et al. (2021), however, we fully expect the MCD12Q1 predictor set to have degraded performance compared to the NLCD predictor set due to the differences in resolution. We included the MCD12Q1 predictor set to highlight the importance of selecting datasets with appropriate resolution for the given modeling task. For large scale studies, it may be impractical to use datasets with such high resolution (e.g., NLCD), but for more localized studies, such as ours, the higher resolution data may lead to improved model performance.

Proportional summaries of classified land cover data are composed of the proportions of all land cover classes present in the given region. To help determine if any changes in model performance across the classified and unclassified predictor sets are due to the compositional information intrinsic to proportional summaries, we discretized the NLCD summaries with binary indicators to represent all present classes in the buffered regions. The discretized NLCD summaries only indicate the presence of land cover classes in place of proportion. While the discretized NLCD summaries do not contain proportional information, they still inform which land cover classes are present and, therefore, still arguably contain more information about the compositional makeup of the region than mean and standard deviation of unclassified imagery.

Experimental Design

Predictor variables

In total, we built 80 models, one for each of the 80 spectral predictor sets (ten spectral datasets with eight season and summary method combinations), for each of the 13 species. For all models we included summaries from the three buffer radii (75, 600, 2400 m) since multi-scale SDMs have been shown to outperform single-scale models (Hallman and Robinson 2020). With the mean values of the raw bands taken over the buffered regions from the summer imagery as our baseline model (i.e., raw bands summer means), we evaluated the effects of adding means from additional seasons, standard deviation of the buffered regions, GLCM texture metrics, and combinations of these summary methods.

In addition to evaluating the unclassified spectral predictor sets, we also included classified imagery in our comparison. We compared models fit using unclassified spectral predictor sets to those fit with a classified spectral predictor set computed according to the best practices of Johnston et al. (2021) (i.e., proportion of land cover classes present in a region surrounding the species record). Johnston et al. (2021) recommends summarizing MCD12Q1 which has 500 × 500 m resolution, which is much larger than the 30 × 30 m resolution of our spectral predictor sets. For a more even comparison of classified to unclassified predictor sets, we included a predictor set derived from 30 × 30 m NLCD data which matches the resolution of our unclassified spectral predictor sets. To investigate if the predictive performance of the classified predictor set had an advantage due to the proportional nature of the summaries, we also included discretized NLCD summaries in our comparison.

We did not perform variable selection as it was unnecessary in this study. Generally, analyses include variable selection for a variety of reasons, including computational considerations, dimensionality reduction, and ease of interpretation, but these motivations did not pertain to our approach. There were no computational considerations, because random forests can accommodate many predictor variables of different types and correlation among input variables does not inhibit the model fitting algorithm. Dimensionality reduction can be a motivation separately from computational issues, for example when it is necessary for all models being compared to have the same number of inputs. This need could arise when fitting and evaluating models on the same training dataset. In such a case, models with more predictor variables have an advantage since they may use additional variables to fit the data more closely, even if the correlations they exploit are spurious (i.e., models with more variables can overfit training data). However, we used spatially distinct training and test sets (described further below) to avoid overfitting; if the models with greater numbers of variables in their predictor sets fit the training data better by exploiting spurious correlations, those correlations would disappear in the test data, resulting in lower predictive performance. In our study design, if models with larger predictor sets perform better on the test data, then they reflect additional information about the species-environment relationship that generalizes to the test data. Note that this is true not just for random forests but also for other modeling approaches. Additionally, variable selection may be used to aid model interpretation by reducing correlation among input variables, which is a major hurdle for determining variable importance. Indeed, remotely sensed inputs are generally highly correlated (Zimmermann et al. 2007), but this is not an issue for conclusions drawn from the predictive power of random forests, as long as the correlation structure remains constant across training and test sets (Dormann et al. 2012).

Species distribution models

We compared the performance of the spectral predictor sets by predicting species occurrences with random forest models. For each species analyzed, we fit the random forest models to predict detection versus non-detection at every count location. Random forests can fit nonlinear relationships between predictors and the response variables automatically (Cutler et al. 2007). This flexibility allowed us to compare the overall performance of the different predictor sets without committing to particular functional forms (e.g., linear) of their effects on the response. Random forests have only two tuning parameters and since our preliminary analyses indicated that our models were not sensitive to these parameters, as is the common case (Breiman 2001; Genuer et al. 2008), we set the number of variables to consider at each split to the default of the square root of the number of predictor variables and the number of trees to fit to be 5000. All of the species we analyzed had more non-detections than detections, and some had very few detections, resulting in substantial class imbalance (Table 1). To address this issue, we used balanced random forests (Chen et al. 2004), which select an equal number of detections and non-detections in the bootstrap sample drawn for each tree by down-sampling the majority class. Balanced random forests is a method suggested by Johnston et al. (2021) for handling class imbalance. We fit all random forest models in R version 3.6.0 (R Core Team 2019) with package 'randomForest' (Liaw and Wiener 2002) and set parameter sampsize to create balanced trees.

Performance estimates computed on spatial data may be biased by spatial autocorrelation when training and test points are close to one another (Roberts et al. 2017). To address this, we split the data into ten spatially distinct folds using the R package 'blockCV' (Valavi et al. 2019). We imposed a 10 × 10 km grid over the study region, numbered the grid cells, and let blockCV randomly assign each cell to one of the ten folds. This process was repeated 100 times and the best assignment of grid cells to folds was kept, as determined by blockCV (evaluated by the most uniform spread of presences and absences per fold) (Valavi et al. 2019). The 10 folds were fixed across all models for a species to ensure that models for each variable set were built and tested on the same data. We then evaluated models with 10-fold cross validation. With this method, one spatial fold is withheld from the training data and all model evaluation is conduction on the withheld fold. The process is repeated ten times to obtain an evaluation of model performance based on all data. Since models are never evaluated with the same data on which they are trained, test data retain a degree of independence.

With our 10-fold cross validation scheme, we evaluated model performance with the area under the receiver operating characteristic curve (AUC) and computed 95% DeLong confidence intervals using the R package 'pROC' (Robin et al. 2011). We chose AUC to avoid the subjective, potentially model- and species-specific process of selecting a classification threshold. While issues with the AUC's ability to assess absolute model performance have been noted in the literature (Lobo et al. 2008), AUC is appropriate for our model comparison task. To assess whether the AUCs were overly optimistic, as can be the case with highly imbalanced data (Davis and Goadrich 2006), we randomly down-sampled non-detections in the independent test set to obtain an equal number of detections and non-detections. Having an equal number of detections and non-detections did not have a substantial impact on the AUCs, so we did not perform any down-sampling when calculating AUCs in the presented results.

Statistical testing

In order to identify which spectral predictor sets performed best across the entire set of species, we compared the performance of the different predictor sets across the group of species with the Friedman analysis of variance test (R's base version) for repeated measures and non-normally distributed data. We controlled for species by calculating the percent difference in AUC from the mean AUC of the predictor sets for each species and subsequently performed all tests on the percent difference in AUC from the species mean AUC. To identify which spectral predictor sets were statistically different, we performed post-hoc analysis with Nemenyi-Tests, R package 'PMCMRplus' (Pohlert 2020) which evaluates pairwise multiple comparisons of mean ranks.

Results

Overall, our models performed well. While Sagebrush Sparrow, a habitat specialist, was the species with the highest performing models with a mean AUC of 0.9666, Western Tanager, a habitat generalist, had the lowest performing models with a mean AUC of 0.6904 across all unclassified spectral predictor sets (Figure 2). Across all species, the raw bands spectral predictor sets were the top performing. Adding seasonal and textural information to the summer means had little impact on the raw-bands and Tasseled-Cap models, but did improved the single-index models (Figure 3, Figure 4). These patterns were consistent across all habitat types and species specialization (Table 5, Figure 2). NLCD, the classified land cover data with the same spatial resolution of Landsat, had equivalent performance to the raw-bands models, whereas MCD12Q1 with its much larger spatial resolution, did not perform nearly as well (Figure 5).

Which index or transformation of the raw bands best predicts species?

Across species, the raw-bands models had the highest AUCs among the summer means spectral predictor sets, with a mean AUC of 0.8990 (Table 5, Figure 3: Summer means). Within individual species, the raw-bands models had the highest AUC for 11 of the 13 species analyzed (Figure 2). Sagebrush Sparrow and Yellow Warbler were better modeled by other spectral predictor sets, but only narrowly, and with the raw-bands models as second best.

Models built with the next highest performing spectral predictor set, the Tasseled Cap transformations, did not statistically differ in performance from the raw-bands models (p-value = 0.9989, Nemenyi post-hoc Friedman). Across species, the Tasseled-Cap models had a mean decrease in AUC from the raw-bands models of only 0.0034.

Across all species, the single-index models had an average 0.0784 decrease in AUC from the raw-bands models. For all but one of the species (and only narrowly), the single-index models were outperformed by the raw-bands models. The highest performing single-index model was the NDVI model which exhibited moderate evidence of being statistically different from the raw-bands models (p-value = 0.0709, Nemenyi post-hoc Friedman) with a mean decrease in AUC from the raw-bands models of 0.0505. The remaining single-index models were all statistically different from the raw-bands models (p-values < 0.0169, Nemenyi post-hoc Friedman). Apart from the NDVI models having the highest average performance across the single-index models, there were no clear patterns as to which indices best predicted species, with different indices producing higher AUCs for different species (Figure 2).

Table 5 Mean rank and standard deviation for the spectral predictor sets calculated across species.

Spectral Predictor	Average Rank	Standard Deviation
Raw bands	1.15	0.38
Tasseled Cap	2.23	0.73
NDVI	4.69	2.02
NBR	5.31	1.75
SAVI	5.92	1.32
NDMI	5.92	2.43
NBR2	6.08	2.75
MSAVI	6.54	1.61
EVI	7.08	2.14

NDSI	10.00	0.00

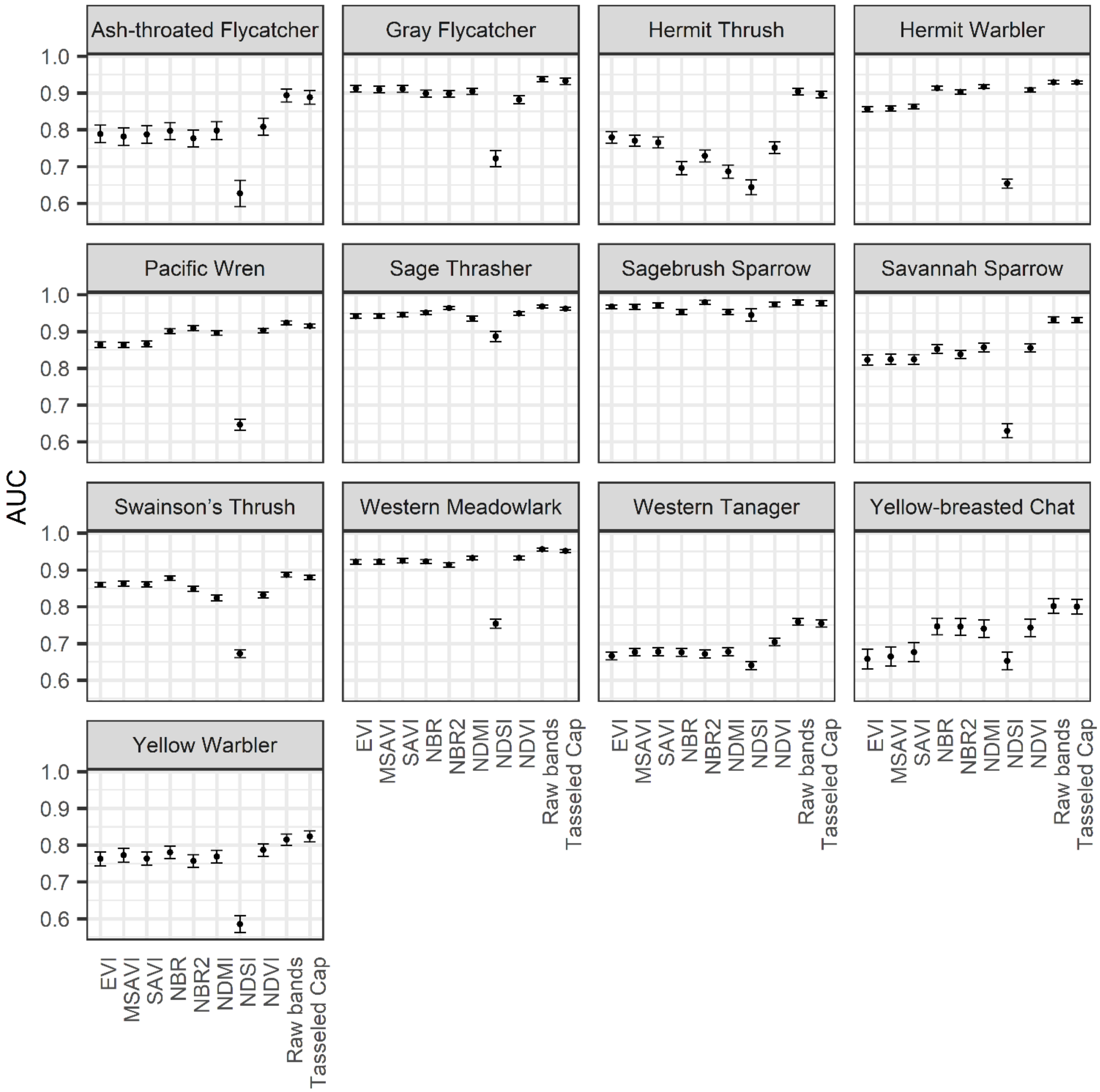

Fig. 2 *Performance of the spectral predictor sets when summarized by their summer means*

How does adding additional seasons to the summer means impact predictive performance?
Adding summaries from spring and fall to the summer spectral predictor sets had a small positive impact on model performance, with an overall average increase in AUC across all models of 0.0445 (Figure 3). The top two performing summer means spectral predictor sets (raw bands and Tasseled Cap) saw a much smaller increase in AUC of 0.0083 compared to the single indices which had an increase in AUC of 0.0536.

How does adding standard deviations and texture metrics to the summer means impact predictive performance?
Across species, inclusion of standard deviations had a small positive impact on model performance with an overall average increase in AUC across all spectral predictor sets of 0.0359 (Figure 3)**.** The top two performing summer means spectral predictor sets (raw bands and Tasseled Cap) saw an increase of 0.0097 in AUC while the single-index predictor sets saw a 0.0424 increase in AUC.

Adding the GLCM texture metrics to the summer means spectral predictor sets also had a small positive impact on model performance, with a mean increase in AUC across all spectral predictor sets of 0.0674 (Figure 3). There was a 0.0127 increase in AUC from the top two summer means spectral predictor sets (raw bands and Tasseled Cap) and a 0.0811 increase in AUC for the single-index predictor sets.

Adding combinations of the additional seasons, standard deviations and texture metrics to the summer means did not have a significant impact on the raw-bands model (Figure 4). For comparison, we present the same analysis for NDVI, a top performing single-index model (Figure 4).

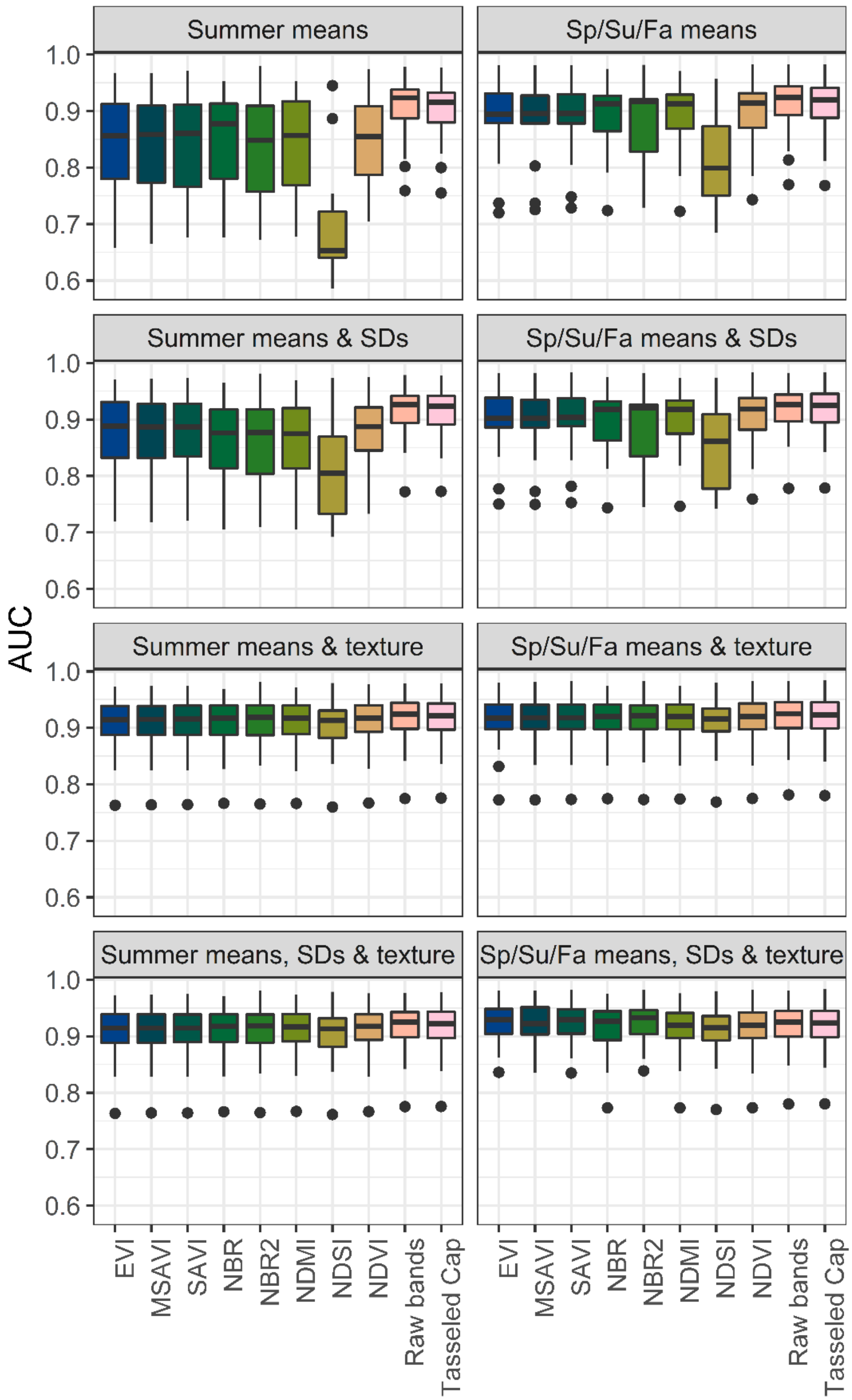

Fig. 3 *Mean AUCs of the spectral predictor sets for each of the summary methods averaged across all 13 species. Black dots indicate outliers that fall outside the whiskers of the box plots.*

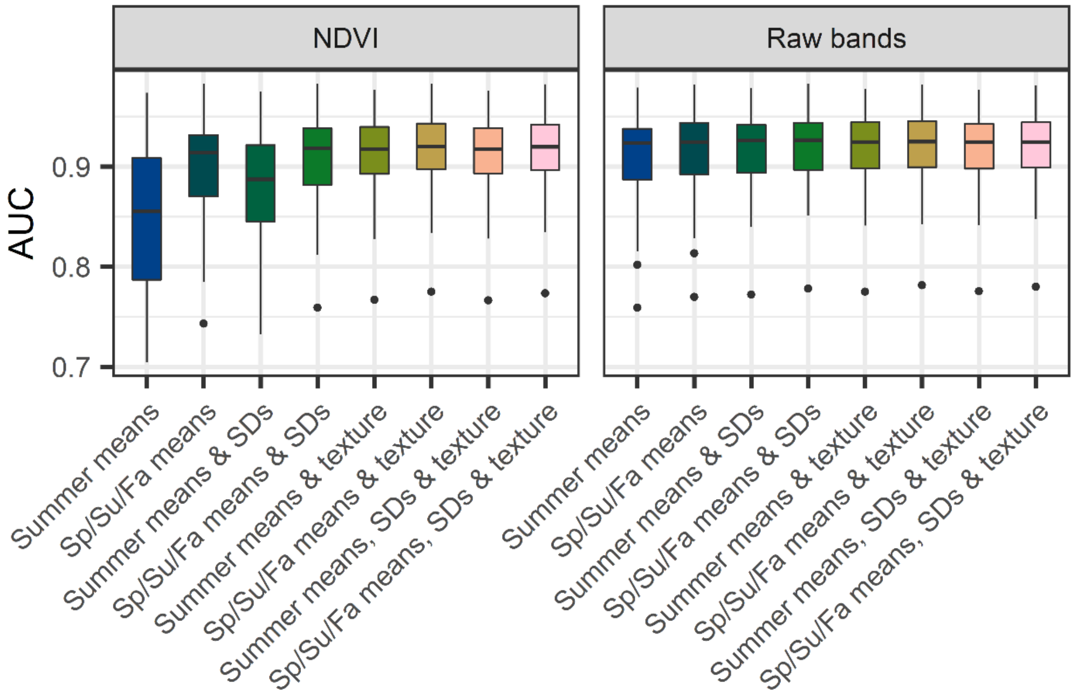

__Fig. 4__ A comparison of the AUCs for all summary methods for NVDI and raw-bands models averaged across all 13 species

How do the unclassified summer means compare to classified remotely sensed predictor sets?

The proportional NLCD summaries had very good performance, with a negligible difference from the raw bands summer means (p-value = 0.9900 Nemenyi post-hoc Friedman; Figure 5). The discretized NLCD summaries did not perform as well as the proportional NLCD summaries, with a mean decrease in AUC of 0.0122 from the raw bands summer means models. Unlike the proportional NLCD summaries, the discretized NLCD summaries were found to be statistically different from the summer means of the raw bands (p-value = 0.0320 Nemenyi post-hoc Friedman; Figure 5). As expected, there was significant evidence that the coarse resolution MCD12Q1 proportional summaries were statistically different from the summer means of the raw bands (p-value = 3.1e-05 Nemenyi post-hoc Friedman; Figure 5) with a 0.0693 decrease in AUC from the raw bands summer means.

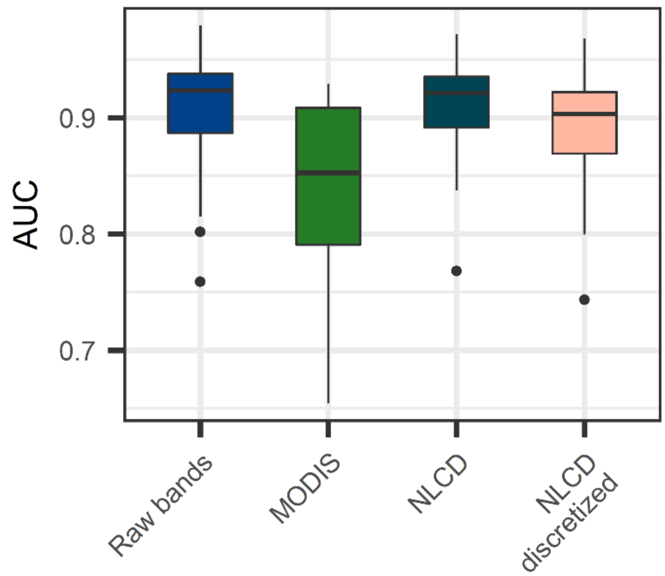

Fig. 5 *Comparison of mean AUCs for classified summaries and the unclassified raw bands summer means. The AUCs are averaged across all 13 species*

Discussion

Our results yielded three important insights regarding models built on unclassified remotely sensed data: 1) raw bands perform better than their summaries, 2) including additional seasons helps single-index models but has little effect on raw-bands or Tasseled-Cap models, and 3) including standard deviations or textural metrics helps single-index models but has little effect on raw-bands or Tasseled-Cap models. Our experimental design protected against overfitting by judging performance on spatially distinct test sets. This strategy is sound for comparing models even with differing numbers of variables, so we can conclude that the performance drop from the raw-bands models (with more variables) to the various reflectance summarizations (with fewer variables) is due to the reduced ability of the latter to characterize the environment. The magnitude of the performance drop speaks to the amount of environmental signal lost. For example, AUCs were greatly reduced from the raw-bands to the NDSI models, because NDSI, an index for characterizing snow, is a substantially inadequate summary for the species in our analysis. In contrast, differences between the raw-bands and the Tasseled-Cap models were negligible, indicating that they are nearly equivalent in their ability to represent relevant signal for predicting species.

1. Raw bands perform better than their summaries

When using mean values alone, models built on raw bands performed consistently better than all other methods of summarization. We saw an insignificant decline in performance following the dimensionality reduction from six raw bands to the four Tasseled Cap transformations and an even larger, and statistically significant, decline in performance with the reduction in dimensions from the raw bands to single indices. It is not surprising that the single indices had degraded performed given that they are computed from a small subset of the raw bands, whereas the Tasseled Cap transformations essentially maintain the principle components of the raw bands. In other words, the single-summary transformations starve the models of environmental information critical for SDMs. Shirley et al. (2013) compared summaries of raw Landsat bands to NDVI and similarly found that raw bands outperformed NDVI in predicting bird distributions. These models do have differing numbers of input variables: six bands and three radii for 18 variables in the raw-bands models, four bands and three radii for 12 variables in the Tasseled-Cap models, and one band and three radii for three variables in the single-index models. If, as in our study, predictive performance is the goal and ML methods that handle many predictor variables are used, we suggest the use of raw bands summarized at multiple radii. Interpreting the effects of specific variables, however, can be difficult with sets of correlated input variables, like variables summarized at multiple radii. Classic approaches such as generalized linear models (GLMs) usually require strict variable selection, but interpretation of the effects of variables is straight forward (e.g., effect sizes and p-values). When the interpretation of variables is the primary objective or other modeling methods are employed, dimensionality reductions may be beneficial. In these cases, we suggest the Tasseled Cap transformations or a single index like NDVI, our highest performing single-index model. If using a single index, however, we recommend including additional summaries, such as standard deviation (Figure 3). Apart from NDVI models having the highest average performance and NDSI models performing consistently poorly across species, nearly all other single-index models performed similarly with a reduction in AUC from the raw band models of about 0.08. Given that NDSI is meant to capture areas of snow well and none of our study species specialize in snowy habitats, it makes sense that it performs the worst of our single-index models. We cannot rule out that NDSI may perform well with species that do specialize in snowy habitats (e.g., Rosy-Finches *Leucosticte* spp.). Methods such as pseudo-scale optimization may be employed to further reduce the number of variables associated with multi-scale models while ensuring that appropriate scales are included (McGarigal et al. 2016).

2. Including additional seasons helps single-index models but has little effect on raw-bands or Tasseled-Cap models

Although the inclusion of additional seasons hardly increased performance in the raw-bands and Tasseled-Cap transformation models, their inclusion did increase the performance of single-index models. Researchers tend to default to spring or "breeding season" environmental data to match the timing of observational data, and perhaps more importantly, because many species are migratory. Twelve of our 13 species are migratory and depart Oregon after their breeding season. Remotely sensed data from winter might therefore be expected to contribute little to explaining distributions. However, the addition of information from other seasons may help to differentiate between habitats whose spectral qualities are similar during a single season (Bino et al. 2008; Senf et al. 2015). For example, deciduous and coniferous forests may have similar spectral qualities during the breeding season, but different spectral qualities following autumnal leaf loss. The seasonal contrast could improve model predictions. Indeed, our findings supported this idea, but increases in model performance were primarily restricted to single-index models (Figure 3). Future studies could consider more complex summaries for quantifying seasonality such as multi-temporal metrics, which could potentially yield greater gains in model performance (Potapov et al. 2019). Habitats and their suitability may be sufficiently described by their unique sets of raw band and Tasseled Cap values, making the inclusion of additional seasons unnecessary.

3. Including standard deviations or textural metrics helps single-index models but has little effect on raw-bands or Tasseled-Cap models

As with additional seasons, although there was little improvement in model performance associated with the inclusion of standard deviations or textural metrics in the raw-bands or Tasseled-Cap models, their inclusion did increase performance in single-index models. Farwell et al. (2020) extracted texture metrics, some identical to those in our study, from two remotely sensed datasets and found the metrics captured several aspects of vegetation heterogeneity that were informative of species richness. Standard deviations or textural metrics add information on the heterogeneity of spectral qualities within a location. This may correspond to the heterogeneity of habitats or the categorization of single habitats with heterogenous spectral qualities (e.g., sparse juniper woodlands). Either way, we might expect an increase in performance. Though our data support this, increases in performance were primarily in single-index models (Figure 3). As with seasons, it may be that the unique combinations of spectral values contained within raw bands and the Tasseled Cap transformations may adequately describe fragmentation and heterogeneity within an area, making the inclusion of standard deviations and textural metrics unnecessary. Based on these findings, we suggest that when using a single index (e.g., NDVI), additional summaries, such as seasonal or textural, should be included.

Classified v. Unclassified data

While several studies have found SDMs built with unclassified data outperform those trained on classified data, our NLCD model had essentially equal performance to our highest performing unclassified model (Figure 5). Cord et al. (2014) compared classified land cover to continuous remotely sensed variables for 30 tree species and found that continuous unclassified data far outperformed classified land cover for predicting distribution patterns. Oeser et al. (2020) found that habitat metrics derived from Landsat Tasseled Cap components and binary snow masks outperformed land cover-based metrics. Given the expected reduction in information associated with transforming continuous raw bands into discrete land cover classes, we were surprised at the high performance of the NLCD models. It is likely, however, the high performance of the NLCD models could be region-specific (i.e., NLCD models may not be comparable to Landsat-based models at a continental scale in which the land cover classes contain more variation in habitat types). Additionally, we suspect that summarizing land cover data by percent cover adds information about land cover composition that is not captured by summaries of unclassified imagery. Though we summarize the central tendency and variance of raw bands with means and standard deviations, these summaries do not necessarily correspond to the quantity of any particular type of habitat. While unclassified data might better characterize environmental differences within a single habitat type, classified data captures the proportions of each habitat type.

The relatively minor loss of performance in our discretized NLCD models, however, indicates that the proportional information may be playing a minor role compared to the added information associated with grouping pixels into discrete land cover classes. By discretizing our NLCD data we removed the proportional information it contained which allowed us to directly examine the importance of proportional information compared to the categorization and occurrence of each class. When we classify habitats from spectral imagery, we inherently add some implicit information on similarities between pixels (e.g., vegetation structure or species composition). This additional information may explain the relatively high performance of models informed by discretized land cover. Further, abundance models are likely more sensitive to information on amount or proportion of habitat than distribution models. When modeling species occurrence, even small areas of suitable habitat can be occupied.

As expected, we found a relatively large loss of performance (0.07 AUC) in models using MCD12Q1 data (Figure 5). While NLCD has the same 30 m resolution as the spectral data, MCD12Q1 is characterized at a 500 m resolution. We anticipated MCD12Q1 would have decreased performance compared to NLCD due to their differences in resolution. As per Johnston et al. (2021), we summarized MCD12Q1 data within a 2500 × 2500 m kernel, which corresponds loosely to a radius of 1250 m. In contrast, we characterized our unclassified remotely sensed data and the NLCD data at three scales: 75, 600, and 2400 m (radii from count location). By including only a single scale, and lower resolution data to begin with, MCD12Q1 data contained less information than NLCD data in this study and may be characterized at a scale too broad to maximize accuracy of predicting local avian occurrences. The differences in model performance between the NLCD and MCD12Q1 predictor sets suggests that when performing localized studies in regions that do not contain high resolution classified data, unclassified data should be considered.

Summaries of classified and unclassified remotely sensed data within buffers differ. Where a mean NDVI value corresponds to some level of vegetation or biomass, it is difficult to translate such a value into real-world management. For example, picturing 55 percent temperate forest within a region is easier to visualize than a mean NDVI value of 0.4957. There are always tradeoffs. One issue with using proportions of land cover is that the number of variables greatly increases (e.g., 16 land cover classes as opposed to a single NDVI value) and this issue is only amplified if researchers are interested in datasets with a greater number of land cover classes. If interested in a small set of specific species, the use of select land cover classes paired with an interpretable modeling method such as GLMs, may be most appropriate. Though we found no decrease in model performance with classified NLCD data, we did not incorporate models with only a subset of pre-determined land cover classes, nor did we test GLM. These methods should be studied in future research.

A main caveat in our study is that our results are based on 13 bird species over the state of Oregon. Although we chose the species to represent a wide diversity of habitats and degrees of specialization, our findings may not apply to organisms that utilize geographic space differently from this set of songbirds or experience different varieties and arrangements of habitats in other geographies. That said, our approach to discern differences in performance could easily be adapted for other species and locations. It is also possible that our results are specific to modelling occurrences and that abundance modeling may reveal different patterns of performance in the environmental predictor sets.

Conclusions

To our knowledge, this is the most extensive study to directly compare the effects of remotely sensed summary methods on SDMs. We analyzed the relative performance of different summary methods for continuous unclassified Landsat data and two classified land cover datasets to help inform which sets of variables are most predictive of bird distributions. Overall, we recommend the use of summer means of the raw bands because they consistently outperformed all other spectral predictor sets and did not require additional seasonal or textural information to achieve their highest performance. However, if fewer variables is imperative, we recommend using the summer mean and standard deviation of NDVI as additional seasons and textural information are important for improving the predictive performance of single indices. Another important, and surprising, finding was the essentially identical performance of the classified NLCD summaries and the raw bands. Contrary to other studies (Cord et al. 2014; Halstead et al. 2019; Oeser et al. 2020), classified summaries did not exhibit a performance decrease compared to the continuous unclassified summaries. While the classified NLCD models achieved equal performance to the raw band models, future work should investigate the source of NLCD’s high performance and evaluate how NLCD-based variables perform in the more challenging task of predicting abundances.

Declarations

Funding
Not applicable

Conflicts of interest/Competing interests
All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethics approval
Not applicable

Consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and material

The datasets generated and analyzed during the current study are available in the A Comparison of Remotely Sensed Environmental Predictors for Avian Distributions repository, https://figshare.com/projects/A_Comparison_of_Remotely_Sensed_Environmental_Predictors_for_Avian_Distributions/94619. The file for a single species has been made available and the remaining species files are available from the corresponding author by request.

Code availability

The code for the current study are available in the Comparison of RS Predictors for Avian Distributions repository, https://github.com/Hutchinson-Lab/Comparison-of-RS-Predictors-for-Avian-Distributions.

Author's contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Laurel Hopkins, Tyler Hallman, and John Kilbride. All authors contributed to writing and editing the manuscript. All authors read and approved the final manuscript.

References

Archer, KJ, Kimes, RV (2008) Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis 52(4): 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015

Baumann, M, Ozdogan, M, Richardson, AD, Radeloff, VC (2017) Phenology from Landsat when data is scarce: Using MODIS and Dynamic Time-Warping to combine multi-year Landsat imagery to derive annual phenology curves. International Journal of Applied Earth Observation and Geoinformation 54: 72-83. https://doi.org/10.1016/j.jag.2016.09.005

Betts, MG, Fahrig, L, Hadley, AS, Halstead, KE, Bowman, J, Robinson, WD, Wiens, JA, Lindenmayer, DB (2014) A species-centered approach for uncovering generalities in organism responses to habitat loss and fragmentation. Ecography 37(6): 517–527. https://doi.org/10.1111/ecog.00740

Bino, G, Levin, N, Darawshi, S, Hal, NVD, Reich-Solomon, A, Kark, S (2008) Accurate prediction of bird species richness patterns in an urban environment using Landsat-derived NDVI and spectral unmixing. International Journal of Remote Sensing 29(13): 3675–3700. https://doi.org/10.1080/01431160701772534

Bradley, BA, Fleishman, E (2008) Can Remote Sensing of Land Cover Improve Species Distribution Modelling? Journal of Biogeography 35(7): 1158–1159

Breiman, L (2001) Random Forests. Machine Learning 45(1): 5–32. https://doi.org/10.1023/A:1010933404324

Buermann, W, Saatchi, S, Smith, TB, Zutta, BR, Chaves, JA, Milá, B, Graham, CH (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. Journal of Biogeography 35(7): 1160–1176. https://doi.org/10.1111/j.1365-2699.2007.01858.x

Chen, C, Liaw, A, Breiman, L (2004) Using random forest to learn imbalanced data. Dept. Statistics, Univ. California, Berkeley (Technical No. 666) University of California, Berkeley

Conners, RW, Trivedi, MM, Harlow, CA (1984) Segmentation of a high-resolution urban scene using texture operators. Computer Vision, Graphics, and Image Processing, 25(3): 273–310

Cord, AF, Meentemeyer, RK, Leitão, PJ, Václavík, T (2013) Modelling species distributions with remote sensing data: Bridging disciplinary perspectives. Journal of Biogeography 40(12): 2226–2227. https://doi.org/10.1111/jbi.12199

Cord, AF, Klein, D, Mora, F, Dech, S (2014) Comparing the suitability of classified land cover data and remote sensing variables for modeling distribution patterns of plants. Ecological Modelling 272: 129–140. https://doi.org/10.1016/j.ecolmodel.2013.09.011

Crist, EP, Cicone, RC (1984) A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap. IEEE Transactions on Geoscience and Remote Sensing, GE-22(3): 256–263. https://doi.org/10.1109/TGRS.1984.350619

Cutler, DR, Edwards, TC, Beard, KH, Cutler, A, Hess, KT, Gibson, J, & Lawler, JJ (2007) Random Forests for Classification in Ecology. Ecology 88(11): 2783–2792

Davis, J, Goadrich, M (2006) The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning 233–240. https://doi.org/10.1145/1143844.1143874

Dewitz, J, (2019) National Land Cover Database (NLCD) 2016 Products: U.S. Geological Survey data release, https://doi.org/10.5066/P96HHBIE

Dormann, CF, Elith, J, Bacher, S, Buchmann, C, Carl, G, Carré, G, Marquéz, JRG, Gruber, B, Lafourcade, B, Leitão, PJ, Münkemüller, T, McClean, C, Osborne, PE, Reineking, B, Schröder, B, Skidmore, AK, Zurell, D, Lautenbach, S (2012) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1): 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Farwell, LS, Gudex-Cross, D, Anise, IE, Bosch, MJ, Olah, AM, Radeloff, VC, Razenkova, E, Rogova, N, Silveira, EMO, Smith, MM, Pidgeon, AM (2020) Satellite image texture captures vegetation heterogeneity and explains patterns of bird richness. Remote Sensing of Environment 253. https://doi.org/10.1016/j.rse.2020.112175

Flood, N (2013) Seasonal Composite Landsat TM/ETM+ Images Using the Medoid (a Multi-Dimensional Median). Remote Sensing 5(12): 6481–6500. https://doi.org/10.3390/rs5126481

Foody GM, Palubinskas, G, Lucas RM, Curran, PJ, Honzak, M (1996) Identifying terrestrial carbon sinks: classification of successional stages in regenerating tropical forest from Landsat TM data. Remote Sensing of Environment 55(3): 205-216. https://doi.org/10.1016/S0034-4257(95)00196-4

Foody, GM (2002) Status of land cover classification accuracy assessment. Remote Sensing of Environment 80(1): 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Foody, GM (2003) Remote sensing of tropical forest environments: Towards the monitoring of environmental resources for sustainable development. International Journal of Remote Sensing 24(20): 4035–4046. https://doi.org/10.1080/0143116031000103853

Friedl, M, Sulla-Menashe, D (2015) MCD12Q1 MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Doi, 10

Genuer, R, Poggi, J-M, Tuleau, C (2008) Random Forests: Some methodological insights (RR-6729). https://hal.inria.fr/inria-00340725/fr/

Gillespie, TW, Foody, GM, Rocchini, D, Giorgi, AP, Saatchi, S (2008) Measuring and modelling biodiversity from space. Progress in Physical Geography: Earth and Environment 32(2): 203–221. https://doi.org/10.1177/0309133308093606

Gorelick, N, Hancher, M, Dixon, M, Ilyushchenko, S, Thau, D, Moore, R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202: 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gottschalk, TK, Huettmann, F, Ehlers, M (2005) Review article: Thirty years of analysing and modelling avian habitat relationships using satellite imagery data: a review. International Journal of Remote Sensing 26(12): 2631–2656. https://doi.org/10.1080/01431160512331338041

Hall, DK, Riggs, GA, Salomonson, VV (1995) Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. Remote Sensing of Environment 54(2): 127–140. https://doi.org/10.1016/0034-4257(95)00137-P

Hallman, TA, Robinson, WD (2020a) Comparing multi- and single-scale species distribution and abundance models built with the boosted regression tree algorithm. Landscape Ecology 35(5): 1161–1174. https://doi.org/10.1007/s10980-020-01007-7

Hallman, TA, Robinson, WD (2020b) Deciphering ecology from statistical artefacts: Competing influence of sample size, prevalence and habitat specialization on species distribution models and how small evaluation datasets can inflate metrics of performance. Diversity and Distributions 26(3): 315–328. http://onlinelibrary.wiley.com/doi/abs/10.1111/ddi.13030

Hallman, TA, Robinson, WD (2021) Building a better baseline to estimate 160 years of avian population change and create historically informed conservation targets. Conservation Biology (in press). https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.13676

Halstead, KE, Alexander, JD, Hadley, AS, Stephens, JL, Yang, Z, Betts, MG (2019) Using a species-centered approach to predict bird community responses to habitat fragmentation. Landscape Ecology 34(8): 1919–1935. https://doi.org/10.1007/s10980-019-00860-5

Haralick, RM., Shanmugam, K, Dinstein, I (1973) Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3(6): 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Haralick, RM, Shanmugam, KS (1974) Combined spectral and spatial processing of ERTS imagery data. Remote Sensing of Environment 3(1): 3–13. https://doi.org/10.1016/0034-4257(74)90033-9

Hardisky, MA, Smart, RM, Klemas, V (1983) Seasonal spectral characteristics and aboveground biomass of the tidal marsh plant, *Spartina-Alterniflora*. Photogrammetric Engineering and Remote Sensing 49: 85–92

He, KS, Bradley, BA, Coord, AF, Rocchini, D, Tuanmu, M-N, Schmidtlein, S, Turner, W, Wegmann, M, Pettorelli, N (2015) Will remote sensing shape the next generation of species distribution models? Remote Sensing in Ecology and Conservation 1(1): 4-18. https://doi.org/10.1002/rse2.7

Hooper, S, Kennedy, RE (2018) A spatial ensemble approach for broad-area mapping of land surface properties. Remote Sensing of Environment 210: 473-489. https://doi.org/10.1016/j.rse.2018.03.032

Huete, AR (1988) A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25(3): 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Huete, A, Didan, K, Leeuwen, WJ, Jacobson, A, Solanos, R, Laing, TD (1999) MODIS vegetation index (MOD13). Algorithm Theoretical Basis Document, 3. /paper/MODIS-VEGETATION-INDEX-(-MOD-13-)-ALGORITHM-BASIS-3-Huete-Didan/2204b55a9ad69e8b69d19e88ad1f0e1f81a5d72b

iNaturalist (2020, October 11) INaturalist. https://www.inaturalist.org

Johnston, A, Hochachka, WM, Strimas-Mackey, ME, Gutierrez, VR, Robinson, OJ, Miller, ET, Auer, T, Kelling, ST, Fink, D (2021) Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. Diversity and Distributions 27(7): 1265 – 1277. https://doi-org.ezproxy.proxy.library.oregonstate.edu/10.1111/ddi.13271

Kennedy, RE, Yang, Z, Cohen, WB (2010) Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. Remote Sensing of Environment 114(12): 2897–2910. https://doi.org/10.1016/j.rse.2010.07.008

Kennedy, RE, Andréfouët, S, Cohen, WB, Gómez, C, Griffiths, P, Hais, M, et al. (2014) Bringing an ecological view of change to Landsat-based remote sensing. Frontiers in Ecology and the Environment 12(6): 339–346.

Kennedy, RE, Yang, Z, Braaten, J, Copass, C, Antonova, N, Jordan, C, Nelson, P (2015) Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. Remote Sensing of Environment 166: 271–285. https://doi.org/10.1016/j.rse.2015.05.005

Kennedy, RE, Yang, Z, Gorelick, N, Braaten, J, Cavalcante, L, Cohen, WB, Healey, S (2018) Implementation of the LandTrendr Algorithm on Google Earth Engine. Remote Sensing 10(5): 691. https://doi.org/10.3390/rs10050691

Kerr, JT, Southwood, TRE, Cihlar, J (2001) Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. Proceedings of the National Academy of Sciences 98(20): 11365–11370. https://doi.org/10.1073/pnas.201398398

Kerr, JT, Ostrovsky, M. (2003) From space to species: ecological applications for remote sensing. Trends in Ecology & Evolution 18(6): 299–305. https://doi.org/10.1016/S0169-5347(03)00071-5

Key, CH, Benson, NC (1999) The Normalized Burn Ratio (NBR): A Landsat TM radiometric measure of burn severity. United States Geological Survey, Northern Rocky Mountain Science Center. Bozeman, MT, USA

Key, CH, Benson, NC (2006) Landscape assessment (LA). In FIREMON: Fire effects monitoring and inventory system. Gen. Tech. Rep. RMRS-GTR-164-CD (Vol. 164, p. LA 1–55). US Department of Agriculture, Forest Service, Rocky Mountain Research Station. https://www.fs.usda.gov/treesearch/pubs/24042/24042

Krishnaswamy, J, Bawa, KS, Ganeshaiah, KN, Kiran, MC (2009) Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. Remote Sensing of Environment 113(4): 857–867. https://doi.org/10.1016/j.rse.2008.12.011

Liaw, A, Wiener, A (2002) Classification and Regression by randomForest. R News 2(3), 18–22

Lobo JM, Jiménez-Valverd, A, Real, R (2008) AUC: A misleading meausre of the performance of predictive distribution models. Global Ecology and Biogeography 17(2): 145-151. https://doi-org.ezproxy.proxy.library.oregonstate.edu/10.1111/j.1466-8238.2007.00358.x

Marshall, DB, Hunter, MG, Contreras, AL (2003) Birds of Oregon: A General Reference. Oregon State University Press

Mazerolle, MJ, Villard, M-A (1999) Patch characteristics and landscape context as predictors of species presence and abundance: a review. Ecoscience 6:117–124. https://doi.org/10.1080/11956860.1999.11952204

McGarigal, K, Wan, HY, Zeller, KA, Timm, BC, Cushman, SA (2016) Multi-scale habitat selection modeling: A review and outlook. Landscape Ecology 31(6): 1161–1175. https://doi.org/10.1007/s10980-016-0374-x

Meddens, AJH, Hicke, JA, Vierling, LA, Hudak, AT (2013) Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery. Remote Sensing of Environment 132: 49-58. https://doi.org/10.1016/j.rse.2013.01.002

Meigs, GW, Dunn, CJ, Parks, SA, Krawchuk, MA (2020) Influence of topography and fuels on fire refugia probability under varying fire weather conditions in forests of the Pacific Northwest, USA. Canadian Journal of Forest Research 50(7): 636–647. https://doi.org/10.1139/cjfr-2019-0406

Mulla, DJ (2013) Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering 114(4): 358–371. https://doi.org/10.1016/j.biosystemseng.2012.08.009

Niittynen, P, Heikkinen, R, Luoto, M (2018) Snow cover is a neglected driver of Artic biodiversity loss. Nature Climate Change, 8, 997-1001. https://doi.org/10.1038/s41558-018-0311-x

Oeser, J, Heurich, M, Senf, C, Pflugmacher, D, Belotti, E, Kuemmerle, T (2020) Habitat metrics based on multi-temporal Landsat imagery for mapping large mammal habitat. Remote Sensing in Ecology and Conservation 6(1): 52–69. https://doi.org/10.1002/rse2.122

Osborne, PE, Alonso, JC, Bryant, RG (2001) Modelling landscape-scale habitat use using GIS and remote sensing: A case study with great bustards. Journal of Applied Ecology 38(2): 458–471. https://doi.org/10.1046/j.1365-2664.2001.00604.x

Parviainen, M, Zimmermann, NE, Heikkinen, RK, Luoto, M (2013) Using unclassified continuous remote sensing data to improve distribution models of red-listed plant species. Biodiversity Conservation. 22: 1731–1754. https://doi.org/10.1007/s10531-013-0509-1

Pflugmacher, D, Krankina, ON, Cohen, WB, Friedl, MA, Sulla-Menashe, D, Kennedy, RE, Nelson, P, Loboda, TV, Kuemmerle, T, Dyukarev, E, Elsakov, V, Kharuk, VI (2011) Comparison and assessment of coarse resolution land cover maps for Northern Eurasia. Remote Sensing of Environment 115(12): 3539–3553. https://doi.org/10.1016/j.rse.2011.08.016

Pflugmacher, D, Cohen, WB, Kennedy, RE (2012) Using Landsat-derived disturbance history (1972-2010) to predict current forest structure. Remote Sensing of Environment 122: 146–165. https://doi.org/10.1016/j.rse.2011.09.025

Pohlert, T (2020) PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended. https://CRAN.R-project.org/package=PMCMRplus

Potapov, P, Tyukavina, A, Turubanova, S, Talero, Y, Hernandez-Serna, A, Hansen, MC, Saah, D, Tenneson, K, Poortinga, A, Aekakkararungroj, A, Chishtie, F, Towashiraporn, P, Bhandari, B, Aung, KS, Nguyen, QH (2019) Annual continuous fields of woody vegetation structure in the Lower Mekong region from 2000-2017 Landsat time-series. Remote Sensing of Environment 232: 111278. https://doi.org/10.1016/j.rse.2019.111278

Powell, RL, Matzke, N, de Souza, C, Clark, M, Numata, I, Hess, LL, Roberts, DA (2004) Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. Remote Sensing of Environment 90(2): 221–234. https://doi.org/10.1016/j.rse.2003.12.007

Powell, SL, Cohen, WB, Healey, SP, Kennedy, RE, Moisen, GG, Pierce, KB, Ohmann, JL (2010) Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. Remote Sensing of Environment 114(5): 1053–1068. https://doi.org/10.1016/j.rse.2009.12.018

Qi, J, Chehbouni, A, Huete, AR, Kerr, YH, Sorooshian, S (1994) A modified soil adjusted vegetation index. Remote Sensing of Environment, 48(2): 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Randin, CF, Ashcroft, MB, Bolliger J, Cavender-Bares, J, Coops, NC, Dullinger, S, Dirnböck, T, Eckert, S, Ellis, E, Fernández, N, Giuliani, G, Guisan, A, Jetz, W, Joost, S, Karger, D, Lembrechts, J, Lenoir, J, Luoto, M, Morin, X, Price, B, Rocchini, D, Schaepman, M, Schmid, B, Verburg, P, Wilson, A, Woodcock, P, Yoccoz, N, Payne, D (2020) Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. Remote Sensing of Environment, 239, 111626. https://doi.org/10.1016/j.rse.2019.111626.

Roberts, DR, Bahn, V, Ciuti, S, Boyce, MS, Elith, J, Guillera-Arroita, G, Hauenstein, S, Lahoz-Monfort, JJ, Schröder, B, Thuiller, W, Warton, DI, Wintle, BA, Hartig, F, Dormann, CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40(8): 913–929. https://doi.org/10.1111/ecog.02881

Robin, X, Turck, N, Hainard, A, Tiberti, N, Lisacek, F, Sanchez, J-C, Müller, M (2011) PROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77

Robinson, WD, Hallman, TA, Curtis, JR (2020) Benchmarking the avian diversity of Oregon in an era of rapid change. Northwestern Naturalist 101(3): 180-193. https://doi.org/10.1898/1051-1733-101.3.180

Rouse, J, Haas, R, Schell, J, Deering, D (1974) Monitoring vegetation systems in the Great Plains with ERTS. Third Earth Resources Technology Satellite-1 Symposium 1: Technical Presentations, section A. https://ntrs.nasa.gov/citations/19750020419

Roy, DP, Kovalskyy, V, Zhang, HK, Vermote, EF, Yan, L, Kumar, SS, Egorov, A (2016) Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sensing of Environment 185: 57–70. https://doi.org/10.1016/j.rse.2015.12.024

Senf, C, Leitão, PJ, Pflugmacher, D, van der Linden, S, Hostert, P (2015) Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sensing of Environment 156: 527–536. https://doi.org/10.1016/j.rse.2014.10.018

Seto, KC, Fleishman, E, Fay, JP, Betrus, CJ (2004) Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. International Journal of Remote Sensing 25(20): 4309–4324. https://doi.org/10.1080/0143116042000192358

Shirley, SM, Yang, Z, Hutchinson, RA, Alexander, JD, McGarigal, K, Betts, MG (2013) Species distribution modelling for the people: Unclassified landsat TM imagery predicts bird occurrence at fine resolutions. Diversity and Distributions 19(7): 855–866

Stow, DA, Hope, A, McGuire, D, Verbyla, D, Gamon, J, Huemmrich, F, Houston, S, Racine, C, Sturm, M, Tape, K, Hinzman, L, Yoshikawa, K, Tweedie, C, Noyle, B, Silapaswan, C, Douglas, D, Griffith, B, Jia, G, Epstein, H, Walker, D, Daeschner, S Petersen, A, Zhou, L, Myneni, R (2004) Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems. Remote Sensing of Environment 89(3): 281–308. https://doi.org/10.1016/j.rse.2003.10.018

Strobl, C, Boulesteix, A-L, Kneib, T, Augustin, T, Zeileis, A (2008) Conditional variable importance for random forests. BMC Bioinformatics 9(307). https://doi.org/10.1186/1471-2105-9-307

Sullivan, BL, Wood, CL, Iliff, MJ, Bonney, RE, Fink, D, Kelling, S (2009) eBird: A citizen-based bird observation network in the biological sciences. Biological Conservation 142: 2282–2292

Thuiller, W, Araújo, MB, Lavorel, S (2004) Do we need land-cover data to model species distributions in Europe? Journal of Biogeography 31(3): 353–361. https://doi.org/10.1046/j.0305-0270.2003.00991.x

U.S. Geological Survey (2016). U.S. Geological Survey Gap Analysis Program, 20160513, GAP/LANDFIRE National Terrestrial Ecosystems 2011. https://doi.org/10.5066/F7ZS2TM0.

Valavi, R, Elith, J, Lahoz-Monfort, JJ, Guillera-Arroita, G (2019) BlockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution 10(22): 225–232

Welch, RM, Sengupta, SK, Chen, DW (1988) Cloud field classification based upon high spatial resolution textural features: 1. Gray level co-occurrence matrix approach. Journal of Geophysical Research: Atmospheres 93(D10): 12663–12681. https://doi.org/10.1029/JD093iD10p12663

White, JC, Wulder, MA, Gómez, C, Stenhouse, G (2011) A history of habitat dynamics: Characterizing 35 years of stand replacing disturbance. Canadian Journal of Remote Sensing 37(2): 234–251. https://doi.org/10.5589/m11-034

Wiens JA, Milne BT (1989) Scaling of 'landscapes' in landscape ecology, or, landscape ecology from a beetle's perspective. Landscape Ecology 3:87–96

Wilson, EH, Sader, SA (2002) Detection of forest harvest type using multiple dates of Landsat TM imagery. Remote Sensing of Environment 80(3): 385–396. https://doi.org/10.1016/S0034-4257(01)00318-2

Wintle, BA, Elith, J, Potts, JM (2005) Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. Austral Ecology 30(7): 719–738. https://doi.org/10.1111/j.1442-9993.2005.01514.x

Wulder, MA, Loveland, TR, Roy, DP, Crawford, CJ, Masek, JG, Woodcock, CE, et al. (2019) Current status of Landsat program, science, and applications. Remote Sensing of Environment 2*25*: 127–147. https://doi.org/10.1016/j.rse.2019.02.015

Zhu, Z, Woodcock, CE (2012) Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment, 118: 83–94. https://doi.org/10.1016/j.rse.2011.10.028

Zimmermann, NE, Edwards, TC, Moisen, GG, Frescino, TS, Blackard, JA (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. Journal of Applied Ecology 44(5): 1057–1067. https://doi.org/10.1111/j.1365-2664.2007.01348.x