



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Incorporating canopy structure from simulated GEDI lidar into bird species distribution models

Citation for published version:

Burns, P, Clark, M, Salas, L, Hancock, S, Leland, D, Jantz, P, Dubayah, R & Goetz, S 2020, 'Incorporating canopy structure from simulated GEDI lidar into bird species distribution models', *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/ab80ee>

Digital Object Identifier (DOI):

[10.1088/1748-9326/ab80ee](https://doi.org/10.1088/1748-9326/ab80ee)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Environmental Research Letters

Publisher Rights Statement:

This Accepted Manuscript is © 2020 The Author(s). Published by IOP Publishing Ltd.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



ACCEPTED MANUSCRIPT • OPEN ACCESS

Incorporating canopy structure from simulated GEDI lidar into bird species distribution models

To cite this article before publication: Patrick Burns *et al* 2020 *Environ. Res. Lett.* in press <https://doi.org/10.1088/1748-9326/ab80ee>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2020 The Author(s). Published by IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Incorporating canopy structure from simulated GEDI lidar into bird species distribution models

Patrick Burns^{1*}, Matthew Clark², Leonardo Salas³, Steven Hancock⁴, David Leland⁵, Patrick Jantz¹, Ralph Dubayah⁶, Scott J. Goetz¹

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

² Geography, Environment & Planning, Sonoma State University, Rohnert Park, CA, USA

³ Point Blue Conservation Science, Petaluma, CA, USA

⁴ School of Geosciences, The University of Edinburgh, Edinburgh, Scotland

⁵ Madrone Audubon Society, Santa Rosa, CA

⁶ Department of Geographical Sciences, University of Maryland, College Park, MD

* corresponding author (Patrick.Burns@nau.edu), 1295 S Knoles Dr, Flagstaff, AZ 86011

ORCIDiDs

Burns: <https://orcid.org/0000-0001-7971-4177>

Clark: <https://orcid.org/0000-0001-5953-2990>

Salas: <https://orcid.org/0000-0002-8801-548X>

Hancock: <https://orcid.org/0000-0001-5659-6964>

Jantz: <https://orcid.org/0000-0001-5103-2270>

Goetz: <https://orcid.org/0000-0002-6326-4308>

Dubayah: <https://orcid.org/0000-0003-1440-6346>

Leland: <https://orcid.org/0000-0001-7931-2223>

Keywords

Bird species distribution models, GEDI, lidar, canopy structure, machine learning

Abstract

The Global Ecosystem Dynamics Investigation (GEDI) lidar began data acquisition from the International Space Station in March 2019 and is expected to make over 10 billion measurements of canopy structure and topography over two years. Previously, airborne lidar data with limited spatial coverage have been used to examine relationships between forest canopy structure and faunal diversity, most commonly bird species. GEDI's latitudinal coverage will permit these types of analyses at larger spatial extents, over the majority of the Earth's forests, and most importantly in areas where canopy structure is complex and/or poorly understood. In this regional study, we examined the impact that GEDI-derived Canopy Structure variables have on the performance of bird species distribution models (SDMs) in Sonoma County, California. We simulated GEDI waveforms for a two-year period and then interpolated derived Canopy Structure variables to three grid sizes of analysis. In addition to these variables, we also included Phenology, Climate, and other Auxiliary variables to predict the probability of occurrence of 25 common bird species. We used a weighted average ensemble of seven individual machine learning models to make predictions for each species and calculated variable importance. We found that Canopy Structure variables were, on average at our finest resolution of 250-m, the second most important group (32.5%) of predictor variables after Climate variables (35.3%). Canopy Structure variables were most important for predicting probability of occurrence of birds associated with Conifer forest habitat. Regarding spatial analysis scale, we found that finer-scale models more frequently performed better than coarser-scale models, and the importance of Canopy Structure variables was greater at finer spatial resolutions. Overall, GEDI Canopy Structure variables improved SDM performance for at least one spatial resolution for 19 of 25 species and thus show promise for improving models of bird species occurrence and mapping potential habitat.

Social Media Abstract

Canopy structure derived from simulated GEDI lidar is important for predicting bird species occurrence

1. Introduction

Due to the widespread impacts of land cover change, pollution, wildlife exploitation, and climate change on species and ecosystems across the world, biodiversity assessment and monitoring is imperative. The United Nations' latest global assessment report on biodiversity and ecosystem services estimates that the number of species threatened from extinction now reaches an unprecedented 1 million, rates of extinction are accelerating, and current actions are insufficient to reverse these trends (IPBES 2019). Monitoring biodiversity across large areas through rapid, scalable, and accurate remote sensing methods will greatly improve our understanding of species distributions, loss and resilience, drivers of changes, and impacts at multiple ecosystem levels (Pereira et al. 2013; Corbane et al. 2015; He et al. 2015; Rocchini et al. 2015). Proenca and colleagues (2017) reviewed scientific approaches to global biodiversity monitoring, with a focus on GEO BON's Essential Biodiversity Variables, and highlight the use of remote sensing for characterizing ecosystem structure and function. In terrestrial applications, remote sensors provide measurements of vegetation properties, including chemistry, structure and phenology, which are related to species habitat requirements, specifically shelter and food resources. In this regard, many unexplored opportunities with new technologies remain open to investigation. For example, NASA's Global Ecosystem Dynamics Investigation (GEDI; Dubayah et al. 2020) is a light detection and ranging (lidar) sensor that was installed on the International Space Station (ISS) in December 2018. This sensor makes fine-scale measurements of 3D vegetation structure, such as overall canopy height and vertical distributions of plant material. The acquisition of these structural measurements over the majority of the land surface is expected to lead to significant advances in biodiversity monitoring and modeling (Bergen et al. 2009; Goetz et al. 2014; Bakx et al. 2019). In this study, our primary goal is to examine the importance and added value of 3D canopy structure data derived from simulated GEDI lidar waveforms in regional-extent bird species distribution models (SDM).

Bird species have been well-monitored relative to other taxa (Troudet et al. 2017), especially in North America, and recent work incorporating long-term surveys and radar remote sensing suggests a drastic decline of 2.9 billion birds since 1970 (Rosenberg et al. 2019). Similar to other taxa, habitat loss is a major driver of reduction in bird species abundance and range (Jetz et al. 2007). Suitable bird habitat includes necessities like food, water, shelter, and nesting sites. Habitats are characterized at various scales, from broad associations to more descriptive microhabitats. Habitat associations are typically related to land-cover classes, frequently broad vegetation types (e.g. coniferous forest, grassland), which can be extracted from existing

1
2
3 national or global land-cover maps (e.g. MODIS C6; Sulla-Menashe et al. 2019). Habitat
4 associations can be further subdivided into microhabitats (e.g. shaded ground beneath mature
5 canopy) which are characterized by fine-scale topography, climate, and vegetation structure.
6 Measurements of vegetation structure are not widely available at fine scales, but previous work
7 suggests that structural complexity is associated with a greater variety of microhabitats which
8 can lead to higher diversity and abundance due to more opportunities for foraging, shelter, and
9 nesting (Cody 1985, Hunter 1999, Whittaker et al. 2001, Hill et al. 2004).
10
11
12
13
14

15
16 Previous SDM applications have mainly used remote sensing to derive spatially-explicit
17 environmental predictors, such as topography and bioclimatic conditions (Wilson and Jetz
18 2016), or land-cover classes (He et al. 2015; Bradie and Leung 2017). When considering bird
19 diversity, continental-extent distributions are largely modeled with bioclimatic variables (e.g.
20 WorldClim; Fick and Hijmans 2017), which can predict overall physiological constraints, and
21 correlate with broad vegetation patterns that influence habitat (Stralberg et al. 2009; Lawler et
22 al. 2009). However, at regional extents and finer spatial resolution (grain), information on
23 vegetation structure is useful for characterizing microhabitats, particularly in areas with
24 heterogeneous vegetation patterns, for example from topographic variation (Stralberg et al.
25 2009) or land-use/land-cover change (Sohl 2014). Bird species diversity is known to respond to
26 both vegetation structure and floristic diversity (MacArthur and MacArthur 1961; Rotenberry
27 1985; Adams and Matthews 2019), both factors that have been missing in models operating
28 beyond the plot scale. Recent developments in airborne remote sensing have provided missing
29 links at regional to local extents, implying possibilities for broader-extent applications when
30 these technologies are elevated to space.
31
32
33
34
35
36
37
38
39
40

41 Airborne lidar surveys of topography and vegetation structure are becoming more frequent, but
42 are still limited in areal extent due to operational costs. While most airborne lidar sensors
43 typically capture discrete returns from small footprints (< 1 m diameter), airborne and
44 spaceborne waveform sensors capture the full vertical profile of vegetation structure, typically
45 using relatively larger footprints (> 20 m diameter) collected over a wider area (Lim et al., 2003;
46 Anderson et al. 2016; Hancock et al., 2019). Lidar data are already improving our understanding
47 of how vegetation structure influences bird species distributions at multiple spatial scales
48 (Tattoni et al. 2012; Rechsteiner et al. 2017; Carrasco et al. 2019) and several review papers
49 have examined lidar-based studies that consider avian distributions (Vierling et al. 2008; Davies
50 & Asner, 2014; Bakx et al., 2019). In general, current research tends to use airborne lidar and
51
52
53
54
55
56
57
58
59
60

1
2
3 derived metrics related to canopy height, vertical and horizontal variability, and cover, as well as
4 understory density. Most studies report a positive effect of lidar metrics in explaining bird
5 distributions and/or patterns of richness, however the extent and scale of analysis are important
6 factors. Species sense and respond to a range of temporal and spatial scales (Wiens 1989;
7 Levin 1992), and this has important implications for parameterizing and interpreting SDMs.
8 Mayor (2009) and McGarigal et al. (2016) review how hundreds of studies address the issue of
9 scale dependence, finding no consensus. At a minimum, they recommend testing SDMs at a
10 variety of spatial scales, ideally optimizing the scale of each variable (multi-scale optimization).
11
12

13
14
15
16
17 In this study, we focus on implementing GEDI lidar - the first spaceborne lidar mission designed
18 specifically to measure vegetation structure. The instrument is scheduled to collect data for 2
19 years, from 2020 to 2021. Initial GEDI data were released on January 22, 2020, and so our
20 analysis uses 2 years of GEDI data simulated from airborne lidar data. Previous research using
21 NASA's airborne full waveform Land Vegetation and Ice Sensor (LVIS), a precursor to GEDI,
22 showed that lidar metrics were useful for modeling bird species richness, abundance, and
23 habitat quality at local extents (Goetz et al. 2007, 2010). More recent research has used space-
24 based ICESat-1 waveform lidar at broader scales (Goetz et al. 2014), corroborating other
25 studies showing vegetation properties derived from multispectral satellites (MODIS), such as
26 percent tree cover and life form type, along with bioclimatic variables were generally sufficient
27 predictors of breeding bird species richness.
28
29
30
31
32
33
34
35

36 Here we created SDMs for 25 species of birds in Sonoma County (Figure 1) using canopy
37 structure predictor variables as well as additional climate, phenology, and other auxiliary
38 geospatial predictors. Our three objectives are to: (1) generate bird SDMs and determine the
39 importance of individual predictor variables and variable groups for 25 bird species; (2)
40 determine whether including Canopy Structure variables improves SDM performance, while also
41 exploring the impacts of spatial scale; and, (3) map probability of bird species occurrence and
42 associated uncertainty of these predictions.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2. Methods

2.1. Spatial and temporal domain of SDMs

We focus on Sonoma County because there are a relatively large number of geolocated bird species observations and the entire county is covered by airborne lidar data (Fig. 1) which we used for the GEDI simulation. A previous effort, the Sonoma County Breeding Bird Atlas (U.S. Geological Survey, 2019), was based entirely on field observations and mapped occurrence of bird species in Sonoma County at 5-km spatial resolution. We sought to create finer spatial resolution maps showing probability of occurrence by generating three regularly-spaced square grids covering the county at 250-, 500- and 1000-m spatial resolution.

The airborne lidar collection occurred from September to November 2013 and is the primary temporal constraint on our analysis. We chose water years (starting in October and ending in September) 2013 to 2015 to bracket the lidar data collection. Additional remote sensing datasets used for SDMs were temporally-aggregated to the entire three-year period. Due to the limited spatial coverage of species observations from this three-year period, we extended the species observation window to ten years (2006 to 2015).

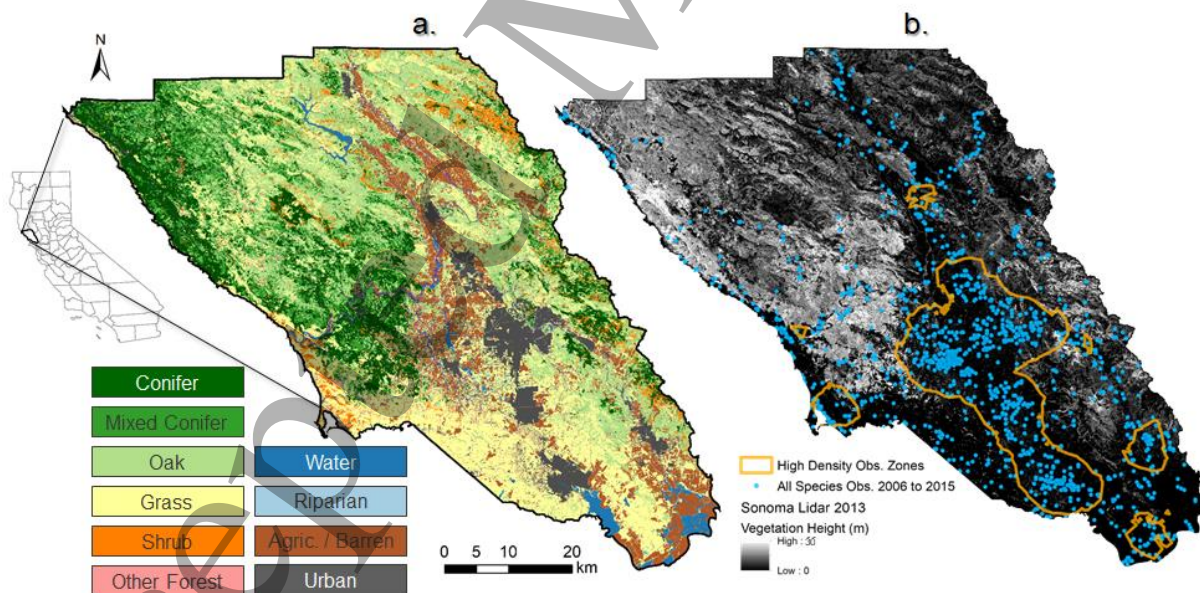


Figure 1. Sonoma County, California, USA dominant vegetation and land cover classes (a.) and vegetation height map (b.) derived from airborne lidar survey in 2013 (sonomavegmap.org). All species observation locations from 2006 to 2015 (blue dots in b.) and corresponding high density observation areas (orange outlines in b.) are shown as well.

2.2. Bird species data

2.2.1. Selecting species of interest

We gathered field observations from point count surveys collected by Point Blue Conservation Science (Fig. 1; Appendix A) from the California Avian Data Center (<http://data.prbo.org/cadc2/>) between 2006 and 2015, along with data from the citizen science observation network eBird (Sullivan et al. 2009), and the Breeding Bird Survey (Sauer et al. 2017). eBird data comprised the majority of records. We used data from all detailed survey events that recorded all species present, and thus from which we could discern detection or non-detection. If presence detections occurred in a given grid cell at a particular resolution then that cell response was classified as presence (value = 1). Grid cell surveys in which the observer did not encounter the species of interest were classified as absence (value = 0). We selected the top 25 species based on number of grid cell presences at 250-m spatial resolution (see Table 1). Main species habitat associations (Conifer, Oak, Shrub, Riparian, Grass, Urban, and Variable; Table 1) were determined from the Sonoma County Breeding Bird Atlas, Audubon Guide to North American Birds (<https://www.audubon.org/bird-guide>) and The Cornell Lab of Ornithology Birds of North America guide (<https://birdsna.org>; Rodewald 2015). A variable habitat association means that the species does not spend a majority of time in a given habitat.

Table 1. Species selected for distribution modeling and their habitat association.

Species Code	Common Name	Pres. Grid Cells at 250-m	Abs. Grid Cells at 250-m	Habitat Association
ACWO	Acorn Woodpecker	189	718	Oak
AMGO	American Goldfinch	114	486	Grass
BEWR	Bewick's Wren	98	615	Variable
BHGR	Black-headed Grosbeak	106	517	Variable
BLPH	Black Phoebe	193	478	Riparian
BRBL	Brewer's Blackbird	172	499	Urban
BUSH	Bushtit	165	579	Oak
CALT	California Towhee	306	438	Shrub
CAQU	California Quail	193	520	Shrub
CBCH	Chestnut-backed Chickadee	127	544	Conifer
DEJU	Dark-eyed Junco	155	589	Conifer

HOFI	House Finch	261	362	Urban
LEGO	Lesser Goldfinch	164	459	Grass
MODO	Mourning Dove	281	741	Urban
NOFL	Northern Flicker	89	624	Conifer
NOMO	Northern Mockingbird	214	808	Oak
NUWO	Nuttall's Woodpecker	152	471	Oak
OATI	Oak Titmouse	181	490	Oak
RWBL	Red-winged Blackbird	175	448	Riparian
SOSP	Song Sparrow	213	610	Riparian
SPTO	Spotted Towhee	137	486	Conifer
STJA	Steller's Jay	119	625	Conifer
WCSP	White-crowned Sparrow	168	432	Shrub
WEBL	Western Bluebird	132	491	Grass
WESJ	Western Scrub-jay	257	414	Oak
	Mean	174	538	
	Std. Deviation	57	108	

2.3. GEDI lidar and canopy structure predictor variables

2.3.1. GEDI Instrument Description

The GEDI lidar sensor measures vertical canopy structure of forests between +/- 51.6 degrees latitude. The instrument emits pulses of energy (1064-nm) which reflect off of various canopy layers and/or the Earth's surface, and the returned energy of each pulse is recorded as a short duration time series. Using the speed of light, ISS orbital geometry, and precise positioning, this time series is geolocated on the Earth's surface and transformed into a waveform - a profile of returned energy discretized into 15-cm vertical height bins for each ~22- to 25-m diameter footprint (Dubayah et al. 2020). Footprints are separated by approximately 60-m along each of the 8 ground tracks and by 600-m between tracks. GEDI, like other lidar systems, is not able to make measurements through clouds.

2.3.2. Canopy Structure predictor variables from simulated GEDI

The primary GEDI data product (L1B) is a geolocated lidar waveform. Variations of the main waveform and waveform-derived metrics (L2; Tang et al. 2012) are output for each footprint.

We simulated two years of GEDI lidar observations from airborne lidar data by using the methods described by Hancock et al. (2019). All simulated L2 variables are based on a gaussian ground-finding algorithm (Hofton et al 2000). The “Canopy Structure” predictor variables used in this study are described in Appendix Table C1. Since GEDI is a sampling instrument we used the “Moving Window Kriging” tool from ArcGIS Pro to produce predicted Canopy Structure metric values at 50-m intervals. We generated the geostatistical model for each Canopy Structure metric using the ArcGIS Pro Geostatistical Wizard and the “Simple” Kriging method. We resampled the 50-m prediction raster to our three different analysis scales using an averaging method. Additional details regarding airborne lidar data processing, GEDI simulation processing and interpolation can be found in Appendix C.a.

2.4. Additional predictor variables

2.4.1. Climate

We selected six “Climate” predictor variables from the California Basin Characterization Model (BCM; Flint et al. 2013) to include in SDM: precipitation (Ppt), average minimum temperature (TMn), average maximum temperature (TMx), potential evapotranspiration (PET), actual evapotranspiration (AET), and climatic water deficit (CWD). The monthly data for water years 2013, 2014 and 2015 were resampled from a native spatial resolution of 270-m to our three analytical spatial scales using an averaging method. Variables for each water year were then grouped and averaged into four three-month periods: October-December (Q1), January-March (Q2), April-June (Q3), and July-September (Q4). Final SDM climate variables were averages of the three water years (2013-2015) for each three-month BCM variable, resulting in 24 total “Climate” variables for each spatial analysis scale.

2.4.2. Auxiliary variables

We used four “Auxiliary” predictor variables as inputs to SDMs: elevation, distance to ocean coast, distance to streets, and distance to streams. Sonoma County has a diverse topography, ranging from Pacific coastlines to mountains greater than 1000-m in elevation. Elevation can be a proximate determinant of a species’ presence (Hof et al. 2012). Elevation rasters were derived from the USGS National Elevation Dataset (NED; 1/3 arcsecond native spatial resolution). In Sonoma County, distance to coast is correlated with a longitudinal pattern of habitats. Closer to the coast are mixed hardwood and conifer forests, such as Coastal redwoods and Douglas fir, with interspersed patches of chaparral, and further inland occur conifer patches, oak woodlands, chaparral and grasslands. Distance to coast was also calculated initially as a 30-m raster from a

1
2
3 coastal vector layer using ArcGIS. Distance to streets was incorporated as a proxy gradient for
4 human influence. We used the Sonoma County streets vector layer to calculate distances in a
5 30-m raster. Distances to streams were incorporated as an additional predictor related to
6 moisture and water availability. Numerous species of birds prefer habitats along or near streams
7 (Rottenborn 1999; McClure et al. 2015). We used the lidar-derived streams vector layer from the
8 Sonoma County Vegetation Mapping and Lidar program (sonomavegmap.org) to calculate
9 distances in a 30-m raster. All auxiliary variables were resampled to the spatial analysis scales
10 (250-m, 500-m, and 1000-m) using an averaging method.
11
12
13
14
15
16

17 2.4.3. Phenology

18
19 Multispectral satellites like the Moderate Resolution Imaging Spectroradiometer (MODIS) are
20 particularly useful for monitoring vegetation presence, vitality, and phenology. We calculated
21 Dynamic Habitat Indices (DHI) Phenological predictor variables from MODIS, most recently
22 described by Hobi et al. (2017) and Radeloff et al. (2019). These indices are measures of
23 vegetation productivity over the course of a year. We were limited to using MOD13 NDVI since
24 it is the only index measured at (approximately) our finest analysis scale. The DHIs we
25 calculated include: 1) DHI sum - the area under the phenological curve of a year, 2) DHI min -
26 the minimum value of the phenological curve of a year, 3) DHI var - the coefficient of variation of
27 the phenological curve of a year, 4) DHI median - the median value of the composite time
28 series, 5) DHI 95p - the 95th percentile value of the composite time series, and 6) DHI seasonal
29 difference - the difference between mean June and mean November NDVI, a potential proxy for
30 deciduousness in this area. The six DHI metrics were calculated at their nominal spatial
31 resolution (232-, 463-, 927-m) and then resampled to our analysis scales. See Appendix C.c. for
32 additional details.
33
34
35
36
37
38
39
40
41
42

43 2.5. Species Distribution Model approach

44 2.5.1. Individual SDMs

45
46 We combined Climate (24), Canopy Structure (20), Phenology (7), and Auxiliary (4) variables for
47 a total of 55 predictors covering Sonoma County at each spatial resolution. We rescaled each
48 predictor variable by subtracting the mean and dividing by the standard deviation. To account
49 for potential multicollinearity we used a variance inflation factor (VIF) analysis to remove the
50 most correlated variables. We used our highest spatial scale scenario (250-m) variable values
51 and set a VIF threshold value greater than or equal to 10 - equivalent to excluding variables that
52 have an R-square of 0.9 or higher when regressed against all other variables. We selected the
53
54
55
56
57
58
59
60

1
2
3 same variables for the other two model scales that use Canopy Structure predictors. This model
4 scenario is referred to as “All with Canopy Structure”. For comparison we ran a scenario
5 referred to as “All without Canopy Structure” which excluded Canopy Structure variables.
6
7

8
9 We ran seven individual SDMs (Random Forests, Support Vector Machine, Boosting, Extreme
10 Gradient Boosting, Neural Network, Net Regularized Generalized Linear Models, and K-Nearest
11 Neighbors) 500 times (bootstraps) for each species, resolution, and model scenario
12 combination. Each bootstrap corresponded to a random sample (80% training, 20% testing) of
13 spatially-thinned species observations (Aiello-Lammens 2015) and associated predictor
14 variables. Cross-validation was used to tune models and prevent overfitting. Additional details
15 related to individual model tuning and bootstrapping can be found in Appendix D. We focused
16 on the area under the receiver operator characteristic curve (AUC; Fielding and Bell 1997) for
17 individual model evaluation. AUC values equal to 0.5 mean that the predicted values are
18 equivalent to random guesses. AUC values greater than 0.5 and up to 1 (perfect) indicate that
19 the model has some level of predictive power.
20
21
22
23
24
25
26
27

28 2.5.2. Variable importance

29 We used a sensitivity analysis method from the R package *rminer* (Cortez 2016; Cortez and
30 Embrechts 2013) to assess variable importance. We selected the data-based sensitivity
31 analysis (DSA) method and associated average absolute deviation from the median (AAD) for
32 measuring importance. We focused on the model scenario “All with Canopy Structure” and
33 assessed variable importance for species both in terms of individual variables and variable
34 groups (Climate, Canopy Structure, Phenology, and Auxiliary). We also assessed variable
35 importance aggregated by habitat association.
36
37
38
39
40
41
42

43 2.5.3. Ensemble SDMs

44 We generated two different types of ensembles, an individual iteration ensemble (IIE) and
45 combined iteration ensemble (CIE). The IIE for each species calculates a weighted average
46 prediction (Marmion et al. 2009) value from up to seven individual model predictions for a single
47 bootstrap (Fig. 2). To create this ensemble we only selected individual models with AUC greater
48 than 0.5 and then calculated a weighted average prediction for each grid cell, where the weights
49 were based on adjusted AUC (AUC score minus 0.5). All prediction averaging was done
50 following a logit transform of the model predictions. Averaged values were then transformed
51 back to probability (0 to 1). Our second ensemble method, CIE, aggregates species model
52
53
54
55
56
57
58
59
60

predictions from individual models across all 500 bootstraps, again weighting the final prediction value by adjusted AUC. The maximum number of individual prediction values was 3500 (7 models * 500 iterations), assuming all individual models had AUC greater than 0.5. We used the weighted average prediction value and associated weighted standard error to display probability of occurrence and associated uncertainty for each species (Fig. 2). All SDM analyses were performed in R (R Core Team 2019). Appendix H lists specific R packages used and associated references.

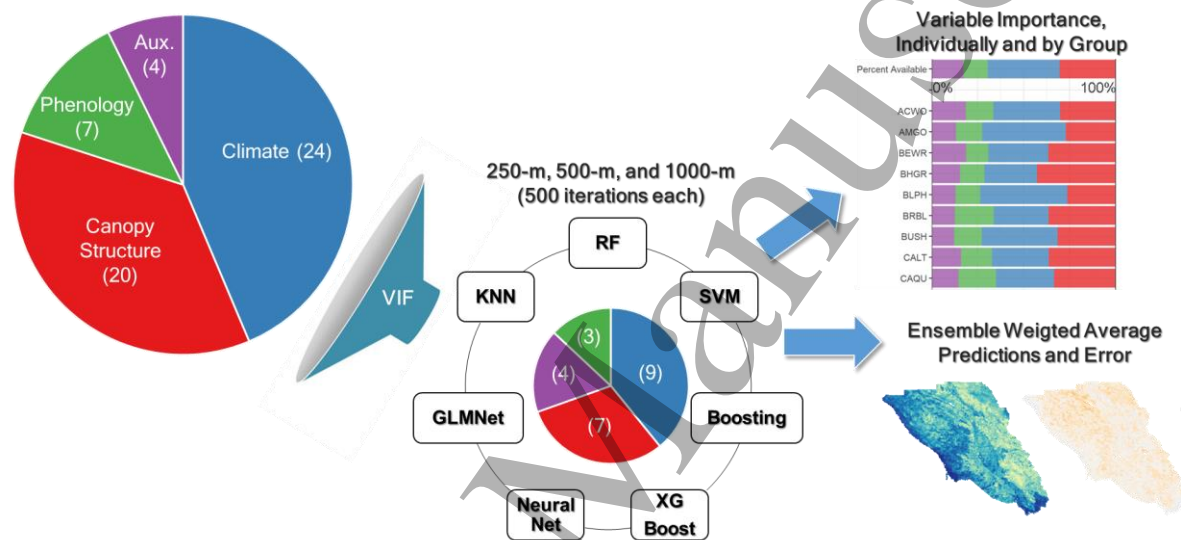


Figure 2. Overview of SDM framework for the “All with Canopy Structure” modelling scenario. VIF reduced the total number of predictor variables to 23. Then for each of the 25 species we used seven different machine learning algorithms at three different spatial scales. We ran 500 bootstraps for each species and scale. We calculated the aggregated variable importance for each species and also produced weighted prediction and uncertainty maps.

2.6. Impact of Canopy Structure variables on model performance

2.6.1. Ensemble SDMs

Differences in predictive accuracy when contrasting model constructs can be highly informative with respect to the importance of a particular predictor for understanding the niche of a species, and for management and decision-making (Austin and van Niel 2010; Mod et al. 2016; Bell and Schlaepfer 2016). However, consistency in identification of important environmental variables varies with the algorithm used and other factors, so caution must be taken to ensure models do not over-fit when tested with and without a variable, as they may include variable interactions that increase model complexity and may not correctly reflect the ecology of the species

1
2
3 (Aguirre-Gutiérrez et al. 2013; Bell and Schlaepfer 2016). This was noted by Araujo and Guisan
4 (2006), who recommend the use of carefully chosen predictors (see also Araujo et al. 2019). We
5 postulated that GEDI variables provide ecologically-meaningful information about vegetation
6 structural complexity that partly determines the niche of several of our study species. We
7 visually compared the performance of SDMs that included and omitted Canopy Structure
8 variables. We expected the algorithms used would adjust model complexity to optimally predict
9 with and without the Canopy Structure variables. However, we also expected Canopy Structure
10 variable importance would still be evident as an increase in performance when they are included
11 in the model.
12
13
14
15
16
17
18

19 Though information from the full range of the species has been noted as very important for SDM
20 performance (Kadmon 2003; Syphard and Franklin 2010; Martínez-Freiría et al. 2015; see
21 review in Engler et al. 2017), the geographic coverage of the airborne lidar dataset precluded us
22 from running models outside Sonoma County. Relatively less important than geographic bias
23 and sample size (Thibaud et al. 2014), spatial autocorrelation will also affect SDM performance
24 and was not included in our models. Since our goal was to evaluate the relative importance of
25 the Canopy Structure variables in SDMs, we focused on building relatively accurate models and
26 comparing among constructs that vary only with respect to the variables used.
27
28
29
30
31
32

33 2.6.2. *Logistic regression analyses*

34 In order to provide additional evidence of the impact of Canopy Structure variables, we used all
35 the variables remaining after VIF to construct an additive logistic regression model (the “All with
36 Canopy Structure” model scenario), and also fitted the “All without Canopy Structure” model
37 scenario. For each bootstrap sample, we then calculated the likelihood ratio test statistic of the
38 full model vs the restricted model (i.e., with vs without Canopy Structure), and report the range
39 of values of the statistic in relation to the value at which it is expected to occur fewer than 5% of
40 the times (the “statistically significant” departure value). Results of this logistic regression
41 analysis are provided for comparison in Appendix E.c. since our main focus is the use of
42 Canopy Structure variables in ensemble SDMs.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

3. Results

3.1. Predictor variables and importance

The VIF method used to remove highly-correlated predictor variables at 250-m spatial resolution reduced the number of predictor variables from 55 to 23. This reduction left 38% of Climate, 35% of Canopy Structure, 100% of Auxiliary and 43% of Phenology predictor variables. The following Climate variables remained: Actual Evapotranspiration (AET) from Q1, Q2, Q3, and Q4, Potential Evapotranspiration (PET) from Q1 and Q2, Precipitation from Q1 (PptQ1), Maximum Temperature from Q1 (TMxQ1), and Minimum Temperature from Q1 (TMnQ1). The following Canopy Structure variables remained: Biomass (BM; referred to as BM_2_1 in Appendix Table C1), Leaf Area Index 0- to 10-m, 10- to 20-m, 20- to 30-m, and 30- to 40-m, Vertical Distribution Ratio Middle (VDRM) and Vertical Distribution Ratio Bottom (VDRB). All Auxiliary variables remained after VIF. The following Phenology variables remained: NDVI Annual 95th Percentile (NDVI_Ann_95p), NDVI Seasonal Difference (NDVI_Seas_Diff), and NDVI Variance (NDVI_Var). The correlation matrix for the remaining variables is shown in Appendix Figure C3.

For all 500 bootstraps we summarized variable importance by predictor variable group for habitat associations (Figure 3) and individual species (Appendix E). We ranked DSA variable importance for each individual bootstrap and then selected the top-5 most important variables from each bootstrap. Figure 3 shows the relative distribution of all top-5 variables by variable group and habitat association. From this perspective at 250-m, Canopy Structure variables most frequently ranked in the top-5 most important variables for Conifer, Shrub, and Urban specialists. The importance of Canopy Structure diminishes at coarser spatial resolutions, such that Climate variables most frequently ranked in the top-5 when considering Shrub and Urban specialists at 1000-m. Across all spatial resolutions and habitat associations, Climate variables most frequently ranked in the top-5, but the relative frequencies generally matched the proportion of Climate variables available relative to the total number of variables. However, Oak, Riparian, and Grass specialists had a higher frequency of Climate variables ranked in the top-5. For all habitat associations, Phenology variables ranked in the top-5 more frequently than would be expected based on proportion of Phenology variables available relative to the total number of variables. Across all habitat associations at 250-m, Auxiliary variables comprised 12.9% of the top-5 (vs. 17.4% expected), Phenology 19.3% (vs. 13.0% expected), Canopy Structure 32.5% (vs. 30.4% expected), and Climate 35.3% (vs. 39.1 % expected). These importance patterns are

corroborated by another perspective in which we summed total importance for each resolution and predictor variable group (Supplementary Figure E2).

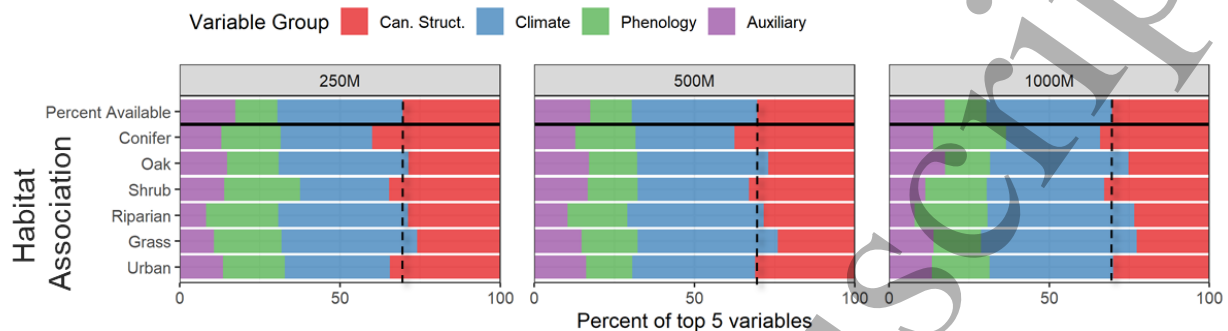


Figure 3. The percent of time that individual variables from different groups were included in the top-5 most important (DSA method) variables across all 500 bootstrap iterations. Percent Available is percentage of variables available from a group (Canopy Structure, Climate, Phenology, Auxiliary) relative to the total number of variables remaining after VIF ($n=23$) and can be thought of as a baseline for comparison if all variables were equally important. The dashed vertical line corresponds to Canopy Structure percent available and is used for assessing the relative importance of this variable group for different habitat associations. We included all models (except XGBoost) which had AUC > 0.5 and used at least 5 variables.

The DSA importance method also provided insight into which individual predictor variables were most important (Appendix Figure E3). Individual Phenology variables were relatively more important than median importance of variables from all models, regardless of habitat association. Canopy Structure variables were usually about as important as the median variable importance. The two vertical distribution ratios (VDRB and VDRM) were more important than other canopy structure variables, with VDRB showing noticeably higher relative importance for all habitat associations. Canopy Structure variables BM, LAI 0- to 10-m, and LAI 10- to 20-m were relatively more important for Conifer specialists, particularly at finer spatial resolution. Climate variables were relatively homogenous in their importance, but TMxQ1 showed noticeable deviation from the median importance value. Precipitation from the first quarter (PptQ1) was another relatively important Climate variable. Importance of Climate variables generally increased going from fine- to coarse-scale.

3.2. Model performance

3.2.1. Comparing performance of different model scenarios

1
2
3 Ensemble models were most frequently the best performing models over all bootstraps
4 (Appendix F.a.). Figure 4 shows notched boxplots corresponding to Ensemble Weighted
5 Average (EWA) AUC values from 500 bootstraps for each species and resolution. These
6 boxplots are useful for visualizing differences among bootstrap population medians. The “All
7 with Canopy Structure” scenario showed median AUC improvements (i.e. non-overlapping
8 notches) for at least one resolution for 19 species: CBCH (all), DEJU (250- and 500-m), NOFL
9 (250- and 500-m), SPTO (250- and 500-m), STJA (all), AMGO (250- and 500-m), BUSH (250-
10 m), NOMO (500-m), NUWO (250- and 500-m), OATI (250-m), WESJ (250-m), RWBL (500-m),
11 SOSP (250-m), CALT (all), CAQU (250-m), WCSP (500-m), BRBL (all), BEWR (250- and 500-
12 m), and BHGR (250-m). Every Conifer specialist showed a median AUC improvement for at
13 least one spatial resolution. Five of the six Oak specialists showed an improvement in
14 performance for at least one spatial resolution. Few of the Grass or Riparian specialists showed
15 improvement after incorporating Canopy Structure variables. All three Shrub specialists showed
16 a median AUC improvement for at least one spatial resolution. Lastly, two of four Urban species
17 showed a median AUC improvement for at least one spatial resolution. Thresholded differences
18 in AUC are shown in Supplementary Table F1. For example, at 250 m, the median AUC value
19 increased by at least 0.02 AUC for 10 of 25 species, while the 95th percentile AUC value
20 increased for 14 of 25 species. The maximum increase in median AUC after incorporating
21 Canopy Structure variables was 0.04 for CBCH.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

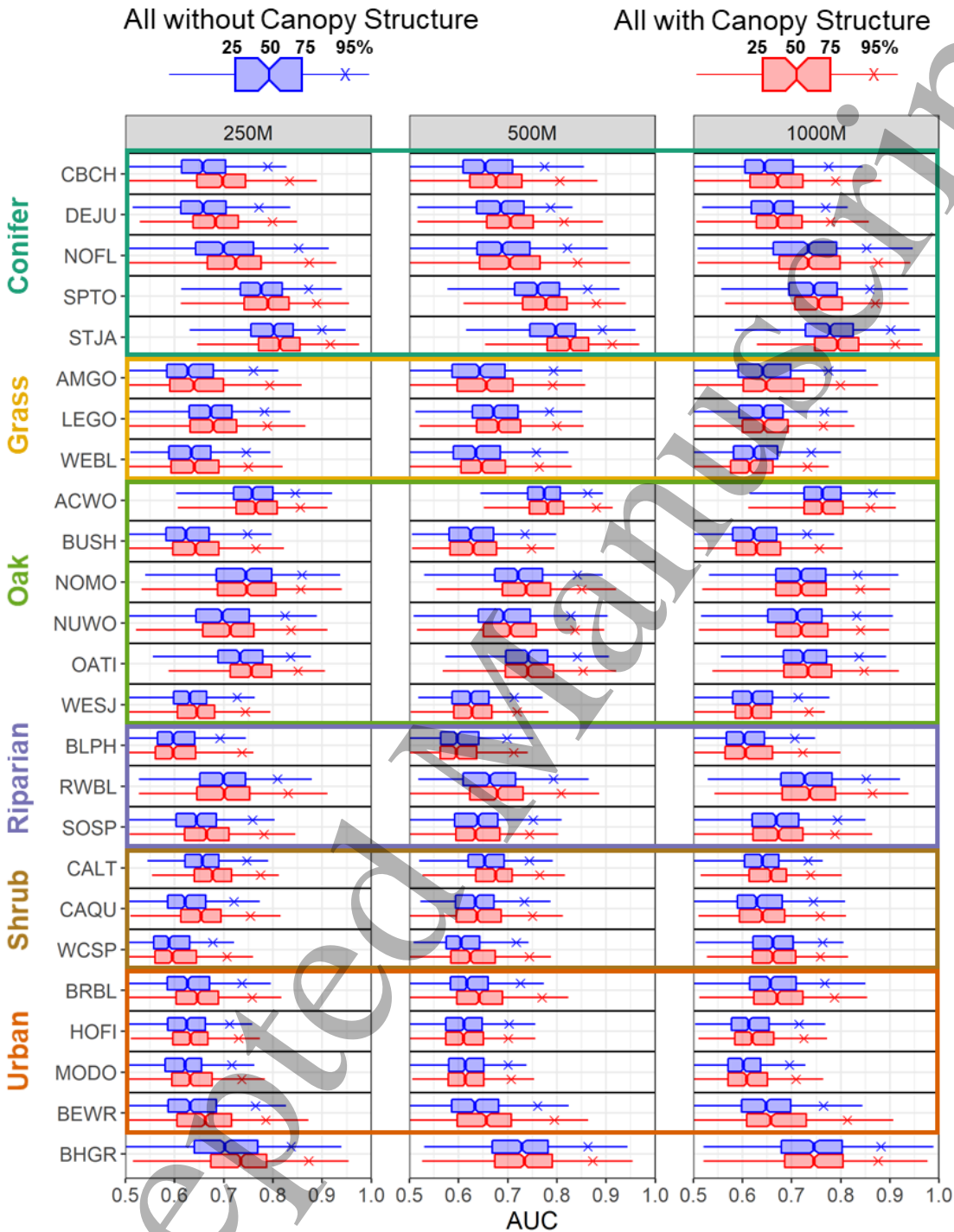


Figure 4. Model performance comparison between two model scenarios - "All with Canopy Structure" vs. "All without Canopy Structure". Each boxplot corresponds to EWA AUC values from 500 bootstraps. Notches that do not overlap can be interpreted as having different median values. "X" points on each boxplot correspond to the 95th percentile value.

3.2.2. Influence of spatial resolution

We summarize “All with Canopy Structure” median model performance by spatial resolution in Supplementary Table F2. Considering all 25 SDMs, most (11) have the highest median AUC value at 250-m. Grass and Riparian habitat specialists are notable exceptions, with all of their highest median AUC values at 500-m and 1000-m, respectively. Conifer, Oak, and Shrub specialists generally have the highest median AUC at 250-m spatial resolution. Supplementary Table F2 also shows the difference between the maximum and minimum AUC medians across all spatial resolutions, providing insight into the scale variability of each SDM. The average scale variability is 0.03 AUC, while the maximum variability is 0.067 AUC. Therefore, the effect of scale on median model performance is similar to the effect of incorporating Canopy Structure variables.

3.3. Ensemble maps

We calculated CIE weighted average predictions from individual models which had AUC values greater than 0.5. Figure 5 shows species distribution maps for birds from a range of habitat associations as predicted by the ensemble weighted average as well as the corresponding uncertainty. The uncertainty estimates include variance due to the model used, the bootstrap sampling, and unexplained variance by the model. Because of the small sample sizes of the bird observation data, the total predicted error is relatively large. All species maps are presented in Appendix G.

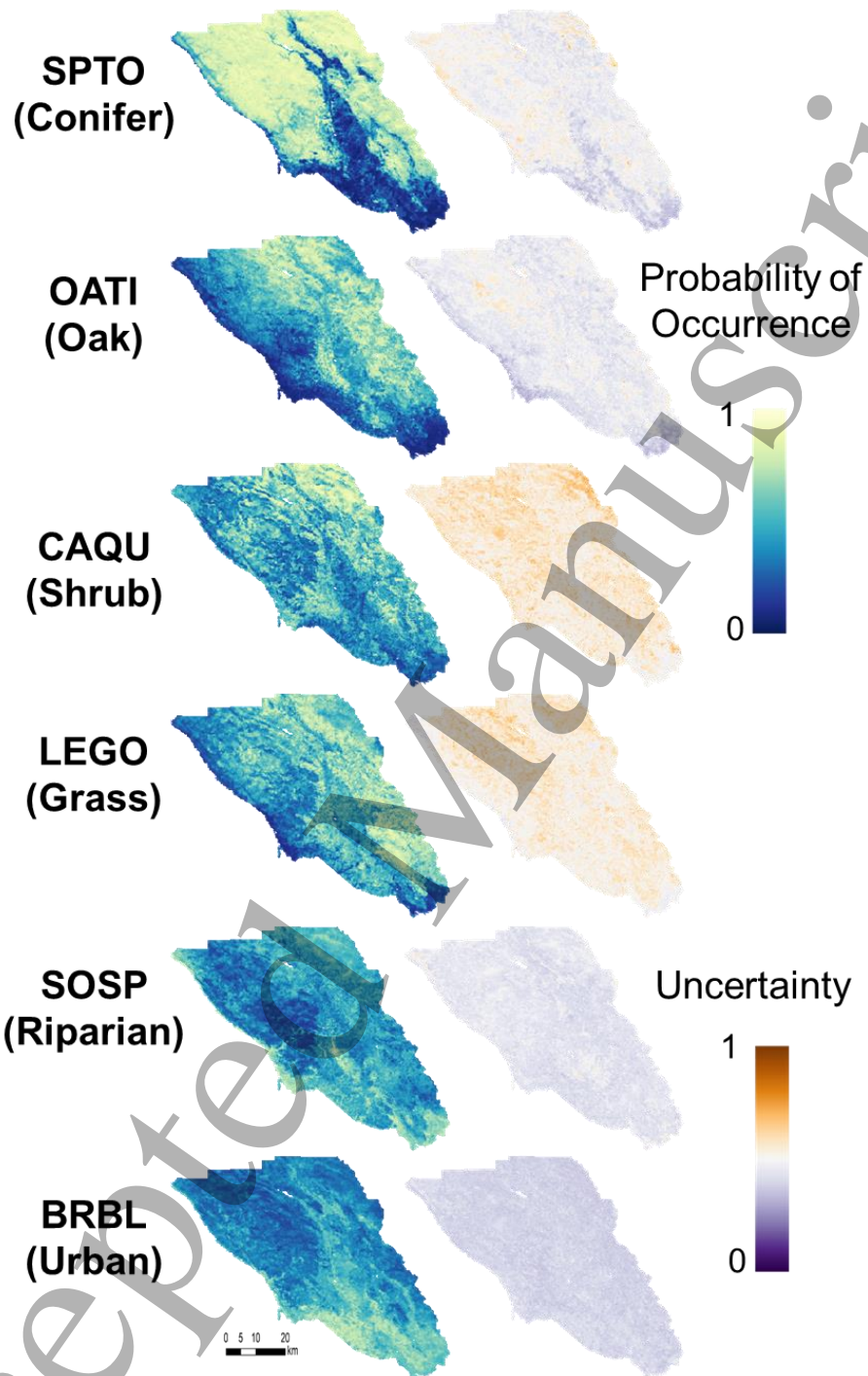


Figure 5. Ensemble maps of weighted average probability of occurrence and associated uncertainty of the mean for one species from each habitat association. Uncertainty of the mean prediction is shown as the range of the standard error confidence interval, that is the upper 1 SE confidence interval minus the lower 1 SE confidence interval.

4. Discussion

Our ability to describe, parameterize, and model species' habitat continues to improve with the addition and refinement of relevant geospatial datasets. Our central objective was to determine the added benefit of Canopy Structure variables from simulated GEDI lidar in SDMs that also include relevant Climate, Phenology, and Auxiliary variables. We quantified the added benefit of these Canopy Structure variables by calculating variable importance of the “All with Canopy Structure” model scenario and by comparing the performance of this model scenario against the “All without Canopy Structure” model scenario using 500 bootstraps. The spatial scale at which SDMs are generated is a very important consideration and impacts our estimates of variable importance and comparative assessment of model performance. For this analysis we selected three practical SDM spatial scales which coincided with readily available and ecologically-meaningful Climate, Phenology, and Auxiliary predictor variables. Below we discuss the variables that remained after the VIF process, their importance, the change in SDM performance after incorporating Canopy Structure variables, current constraints on the analysis, and ideas for building on this work in the future.

4.1. Model variables and importance

Methods to reduce redundant information across predictor variables, such as VIF, are useful for creating models that are parsimonious and interpretable in terms of variable importance. Furthermore, the elimination of certain variables can provide insight regarding unique information content relative to the entire stack of environmental predictors. Considering all variable groups, we found Climate variables to be the most important overall. Maximum temperature was the most important Climate variable for Grass and Oak species, which agrees with previous studies (Howard et al. 2015; Bradie and Leung 2017). Six of nine Climate variables were associated with either potential or actual evapotranspiration (ET), which is related to a combination of temperature, moisture availability, and plant productivity. Hawkins et al. (2003) and Coops et al. (2018) found ET variables to be important for predicting species richness at broad spatial scales. Barbet-Massin and Jetz (2014) showed that annual PET was among the most relevant variables for predicting bird species distributions in the conterminous U.S. The ET variables don't appear to be relatively as important when grouped by habitat association, but still contain some unique information. For example, PETQ2 is one of the most important variables for Bushtit and Mourning Dove.

1
2
3 Seven Canopy Structure variables remained following VIF reduction, and as a group were
4 usually the second most important variable group depending on habitat association. Canopy
5 structure is an important determinant of avian diversity, including the partitioning of niches for
6 species coexistence, changes in microclimate (Zellweger et al. 2019) and refuge from predation
7 (Davies and Asner, 2019). Müller and colleagues (2010) found that 3D structure measured with
8 airborne lidar was the main statistical determinant of bird assemblages in a German temperate
9 mixed forest. MacArthur and MacArthur (1961) first reported the link between bird species
10 diversity and foliage height diversity (FHD). However, neither version of the FHD metrics
11 remained following VIF reduction. This is not necessarily because the variable is without utility,
12 but rather some of its information content varied linearly with other variables that also include
13 information about the vertical distribution of plants, like vertical distribution ratios (VDRB and
14 VDRM), LAI profiles, and NDVI. VDRB was a particularly important variable for all habitat
15 associations. This variable was initially included to characterize how evenly plant material is
16 distributed, as it compares the overall canopy height to the height of 50% of the cumulative
17 waveform signal. In theory, values close to 0.5 indicate an even distribution, while lower values
18 should indicate a higher concentration of plant material in the understory, and vice versa. In
19 actuality, when examining the kriged raster we observe that the variable shows spatial
20 correspondence with tall vegetation or built-up structures, both of which could be useful for
21 perching or nesting. Some species SDMs may find this variable useful for discriminating
22 between canopy structure characteristics. For instance, this was one of the most important
23 variables for California Towhee, which commonly uses dense shrub/scrub areas that are low to
24 the ground.

25
26
27
28
29
30
31
32
33
34
35
36
37
38 Canopy Structure variables were most important for Conifer specialists (Figure 3), suggesting
39 that these species seem to prefer locations with certain Canopy Structure characteristics (i.e.,
40 based on importance metrics, more biomass at the canopy level and denser foliage, or LAI, in
41 the 0- to 10-m and 10- to 20-m strata). This agrees with our observations of these conifer-
42 associated species in Sonoma County. For example, Dark-eyed Junco is often found in areas
43 with complex vegetation structure less than 5-m from the ground, but also forages in mid-
44 canopy. During the breeding season, males often sing from exposed perches near the tops of
45 conifers and snags. Chestnut-backed Chickadee prefers structurally-complex mature conifer
46 forests, as corroborated by the importance of the variable LAI 20- to 30-m, and is often found
47 gleaning small insects and other arthropods from bark and twigs.
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Phenology variables mainly capture temporal cycling of top-of-canopy foliage status, but in
4 some cases (low to moderate forest cover) may contain information about foliage density
5 associated with the near-infrared component of NDVI which is related to the phenomenon of
6 leaf additive reflectance (Jensen 1996). The Phenology variables had the highest top-5
7 importance difference relative to their expected importance (+6.3% overall at 250-m) suggesting
8 that these variables contain particularly useful information for bird SDMs. Visually, the three
9 Phenology variables show some correspondence with broad land cover and vegetation classes
10 (Figure 2), which may provide sufficient predictive information for some SDMs. Canopy
11 Structure variables from GEDI also had a positive top-5 importance difference (+2.1% overall at
12 250-m) for all habitat associations. Canopy Structure variables likely complement Climate and
13 Phenology variables by adding 3D information related to microhabitat structure that is not
14 captured by the other variables we used. This 3D perspective has the greatest benefit for
15 Conifer specialists (+9.5% overall at 250-m).

25 **4.2. SDM performance**

26 **4.2.1. Ensemble model performance perspective**

27 Fewer than half of the ensemble SDMs had median AUC greater than 0.7, a threshold for
28 assessing SDM quality (Pearce and Ferrier 2000; Duan et al. 2014). The overall performance of
29 these 25 SDMs was likely related to the bird observation data and the large distribution range of
30 the individual species (McPherson and Jetz 2007). Although Sonoma County is relatively
31 diverse in terms of climate, topography, and vegetation type, more bird observations spanning a
32 wider range of the environmental predictor variables could lead to better presence/absence
33 separation and thus improved SDM performance. Furthermore, the best performing SDMs were
34 generally associated with Conifer and Oak species, which are relatively more selective and less
35 likely to occur across multiple habitats in Sonoma County. Fairly common species SDMs, such
36 as Mourning Doves, Song Sparrows, and House Finches did not benefit much from the
37 incorporation of Canopy Structure information due to their high prevalence in multiple habitats
38 (Kadmon et al. 2003, Brotons et al. 2004, but see Gavish et al. 2017 for a counter-example).

39
40
41
42
43
44
45
46
47
48
49 Larger observation datasets and more certainty in presence/absence records both help models
50 to more accurately depict the niche of common species (Kadmon et al. 2003). While there are a
51 large number of survey events in Sonoma County (n=10,606), these occurred largely in the
52 same locations -- when aggregated within cells, only a few hundred cells were surveyed (see
53 Table 1). However, because most cells had many survey events, our determination of
54
55
56
57
58
59
60

1
2
3 presence/absence within each cell is probably very accurate, the more so for these common
4 and easy to identify generalist species. It is therefore likely that the relatively poor performance
5 of most models is due to low sample sizes and associated lack of high-density observations
6 from different land-cover classes. These are common limitations associated with citizen science
7 observations, such as those used in this study (Boakes et al. 2010, Beck et al. 2014).
8
9

10
11
12 Although Canopy Structure variables frequently improved median model performance,
13 especially at finer spatial resolution, the improvement in performance was relatively modest
14 considering the number of new variables (and potential amount of new information) added to the
15 models. The various machine learning models used in the ensemble performed well at
16 discerning patterns and optimizing predictions with relatively few predictor variables. Another
17 reason for modest improvements in performance was related to correlation and information-
18 overlap between variables remaining following VIF. Even after using VIF, there was still
19 correlation among the predictor variables (see Appendix Figure C3) suggesting that some
20 canopy structure information could be explained by other variables. For examples we found
21 Pearson correlation coefficients greater than 0.6 when regressing NDVI95p against three
22 Canopy Structure variables (Biomass, LAI 0- to 10-m, and LAI 10- to 20-m).
23
24
25
26
27
28
29
30

31
32 Lastly, spatial scale dependence cannot be ignored in SDM analyses (Wiens 1989; Jackson
33 and Fahrig 2015). In this study, we found an average scale variability of 0.03 AUC
34 (Supplementary Table F2) which is similar to the effect of incorporating Canopy Structure
35 variables in some SDMs. When going from fine to coarse spatial scale, we found Canopy
36 Structure variable importance decreased and incorporation of Canopy Structure variables less
37 frequently improved model performance. Mertes and Jetz (2018) found resampling
38 environmental predictor variables with fine spatial structure decreased the performance of
39 SDMs. Curry et al. (2018) found that spatially-explicit ensemble models with finer-scale
40 predictor variables consistently performed better at predicting occurrence for 3 of 11 grassland
41 bird species. Our results, and those from others, are related to the spatial structure and
42 autocorrelation of Canopy Structure variables (Mertes and Jetz 2018). We did not explicitly
43 factor in spatial autocorrelation of environmental predictor variables into SDMs, but
44 semivariograms generated for each Canopy Structure variable confirmed the spatial structure
45 (semivariogram range) of each variable was generally within 250- to 1000-m (Supplementary
46 Table C2). The majority of semivariograms generated from GEDI simulation subsets within
47 Conifer and Oak forests have ranges below 500-m. Additional SDM performance improvements
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 may be possible through exploration of multi-scale optimization (e.g. Wan et al. 2017; Stevens
4 and Conway 2019), especially when considering even finer-scale vertical (Seavy et al. 2009;
5 Gastón et al. 2017) and horizontal (Huang et al. 2014; Carrasco et al. 2019) Canopy Structure
6 variables.
7
8
9

10 11 **4.3. Constraints on the current analysis**

12 The primary limitations of this study are the quantity and geographic distribution of bird
13 observations, as well as our ability to represent canopy structure at optimal scales using
14 simulated spaceborne lidar data. The biggest limitation associated with the bird species
15 observation data is a lack of high-density, evenly-distributed spatial observations. As shown in
16 Figure 1, most observations occurred in urban/suburban areas. Since we are primarily
17 interested in how forest Canopy Structure variables improve SDMs, this was not the ideal
18 network of observations because relatively few observations were distant from roads and inside
19 of intact forests. The impact of these Canopy Structure variables is likely more difficult to resolve
20 when urban features are mixed with tree canopies at the grid cell resolutions used in this
21 analysis. We would expect to receive a more consistent signal from a “natural” forest canopy
22 compared with a suburban or urban-wildland interface canopy interspersed with built-up
23 structures.
24
25
26
27
28
29
30
31
32

33 Other limitations are associated with our ability to represent canopy structure using simulated
34 spaceborne lidar data. First, the simulated lidar data are hypothetical geospatial observations -
35 actual mission footprint locations will vary as a function of orbital geometry, which cannot be
36 predicted exactly, and local cloud cover patterns. Another major limitation concerns interpolation
37 from footprints (L2 products) to grids (L3 products). There are numerous options for
38 interpolating point data to continuous grids depending on the nature of the dataset. As a starting
39 point, we used a kriging methodology similar to the one outlined in the GEDI L3 Algorithm
40 Theoretical Basis Document (Dubayah et al.2020). Over the course of two years, GEDI is
41 expected to provide sufficient coverage for generating L3 products at 1000-m spatial resolution.
42 However, coverage gaps will still be present in 1000-m spatial resolution grids and would be
43 even more prevalent at the finer spatial resolutions used in this study (see Appendix Figure C2).
44 For natural and continuous forest types, we expect this kriging method to be a good option for
45 interpolation as neighboring grid cells tend to have similar geostatistical characteristics. For
46 fragmented or urban/suburban forests, this method may not be as effective. More advanced
47 interpolation methods such as co-kriging (Tsui et al., 2013) or fusion (Swatantran et al 2012)
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 with auxiliary data with continuous spatial coverage, like Sentinel-1, Sentinel-2 or Landsat
4 satellite imagery, may prove to be better options for interpolating large areas or filling in
5 sparsely-sampled grid cells.
6
7

8 9 **4.4. Future directions**

10 We expect canopy structure measurements from GEDI will be informative for modelling habitat
11 for a variety of forest-dependent taxa. Our modeling approach is applicable to larger spatial
12 extents and taxa, but one important consideration would be the source of climate data, as BCM
13 is a California-wide climate product with an unusually fine spatial resolution (270-m). Global
14 climate variables from WorldClim (Fick and Hijmans 2017), which have approximately 1000-m
15 spatial resolution, are more commonly used for larger SDM extents outside of the conterminous
16 U.S. We note that this resolution matches the planned GEDI L3 products. It will be necessary to
17 test the utility of these global climate variables and GEDI L3 metrics in SDMs at their native
18 resolution. However, for some species the tradeoff in model performance as a function of spatial
19 scale may preclude the use of relatively coarse vegetation structure and/or climate data
20 (Manzoor et al. 2018). Alternatively, shifting to more of a microhabitat focus, recent studies have
21 proposed frameworks for dynamically modeling microclimate using fine-scale topography and
22 vegetation structure datasets derived from remote sensing (Kearney et al. 2019; Lembrechts et
23 al. 2019). Using these finer spatial and temporal scale climate datasets in conjunction with
24 vegetation structure measurements will improve our understanding of the mechanistic interplay
25 between the two, as well as how species and ecosystems respond to changes in climate
26 (Zellweger et al. 2019).
27
28
29
30
31
32
33
34
35
36
37

38
39 It will also be beneficial to explore data and methods for increasing the quantity and coverage of
40 canopy structure measurements. GEDI is scheduled for two years of data collection on the ISS
41 but a different type of spaceborne lidar data, being collected simultaneously from ICESat-2, may
42 have some utility in filling GEDI canopy structure coverage gaps. The Advanced Topographic
43 Laser Altimeter System (ATLAS) instrument on-board ICESat-2 uses photon counting
44 technology which has limitations in dense vegetation, especially during daytime observations
45 (Popescu et al. 2018). Although vegetation measurement is not the primary focus of the ICESat-
46 2 mission, we still expect some useful canopy structure observations in forests of sparse to
47 moderate canopy cover (Neuenschwander et al. 2019).
48
49
50
51
52
53
54
55
56
57
58
59
60

5. Conclusions

We incorporated Canopy Structure variables derived from simulated GEDI lidar into 25 bird SDMs in Sonoma County, CA. We found Canopy Structure variables were generally the second most important group of variables across a variety of habitats. Canopy Structure was most important for Conifer and Shrub habitat associations. The incorporation of Canopy Structure variables into ensemble SDMs resulted in performance improvements for the majority of species across all spatial analysis resolutions. More SDMs showed higher, albeit modest, performance improvements at the finest spatial analysis scale suggesting that Canopy Structure variables are more beneficial at smaller scales which are more in line with microhabitats. Although this study covers a regional spatial extent and limited number of species, it demonstrates that Canopy Structure variables derived from spaceborne lidar have the potential to improve our ability to map species distributions.

Our understanding of global forest canopy structure will grow by orders of magnitude over the next decade as a result of the GEDI and ICESat-2 spaceborne lidar missions. We encourage future SDM efforts to explore the utility of operational lidar products from these sensors across natural and anthropogenic landscapes. Future SDM studies may also benefit from fusion of spaceborne lidar with other remote sensing data as this combination can provide more continuous spatial coverage across a broader range of spatial scales. SDMs generated at even finer spatial scales may benefit more from lidar-derived Canopy Structure variables, resulting in more detailed map products for land management and conservation of species habitat.

6. Acknowledgements

Airborne lidar data were provided by the University of Maryland and the Sonoma County Vegetation Mapping and Lidar Program under grant NNX13AP69G from NASA's Carbon Monitoring System (Dubayah, PI). We acknowledge Curt Joyce at Quantum Spatial Inc. who helped to acquire the airborne lidar data. Curt , a leader in outdoor, remote sensing, and environmental communities, passed away in 2014. Xiaoli Sun conducted the link margin analysis we used to set the noise levels in the GEDI simulation. Scott Luthcke produced orbital simulations and gave expert suggestions regarding expected data quality/quantity.

Bird observations were largely collected and organized by the invaluable contribution of a number of citizen scientists, through eBird (Sullivan et al. 2009) and the BBS programs.

We used Northern Arizona University's high performance computing system (Monsoon) to run all species distribution models. Maxwell Gilbert, Emma Forester, and Mariana Palacio Salazar assisted with pre-processing of predictor variable datasets. Sean Mahoney provided contextual knowledge about bird species of interest. Wendy Schackwitz provided a helpful review.

We acknowledge funding support from NASA's Citizen Science for Earth Systems Program (Cooperative Agreement 80NSSC18M0107 to MC) and the GEDI Science Definition Team (NNL15AA03C to SJG).

Lastly, the manuscript benefited substantially from two anonymous reviews.

7. Code and Data Availability

The data that support the findings of this study are available upon request from the authors.

Relevant model input data and code are stored on GitHub:

https://github.com/leosalas/Soundscapes2Landscapes/tree/master/GEDI_Bird_SDM

8. References

- + Adams, B.T. and Matthews, S.N., 2019. Diverse temperate forest bird assemblages demonstrate closer correspondence to plant species composition than vegetation structure. *Ecography*, 42(10), pp.1752-1764.
- +Aguirre-Gutiérrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M. and Biesmeijer, J.C., 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria—Dutch hoverflies as a case study. *PloS one*, 8(5), p.e63708.
- +Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B. and O'Hara, R.B., 2019. Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), p.eaat4858.
- +Araujo, M.B. and Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *Journal of biogeography*, 33(10), pp.1677-1688.
- +Austin, M.P. and Van Niel, K.P., 2011. Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, 38(1), pp.1-8.
- +Bae, S., Müller, J., Lee, D., Vierling, K.T., Vogeler, J.C., Vierling, L.A., Hudak, A.T., Latifi, H. and Thorn, S., 2018. Taxonomic, functional, and phylogenetic diversity of bird assemblages are oppositely associated to productivity and heterogeneity in temperate forests. *Remote sensing of environment*, 215, pp.145-156.
- +Bakx, T.R., Koma, Z., Seijmonsbergen, A.C. and Kissling, W.D., 2019. Use and categorization of Light Detection and Ranging vegetation metrics in avian diversity and species distribution research. *Diversity and Distributions*.
- +Barbet- Massin, M. and Jetz, W., 2014. A 40- year, continent- wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions*, 20(11), pp.1285-1295.
- +Beck, J., Böller, M., Erhardt, A. and Schwanghart, W., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, pp.10-15.
- +Bell, D.M. and Schlaepfer, D.R., 2016. On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological modelling*, 330, pp.50-59.
- +Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G. and Radeloff, V.C., 2009. Remote sensing of vegetation 3- D structure for biodiversity and habitat: review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences*, 114(G2).

- 1
2
3 +Boakes, E.H., McGowan, P.J., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K. and
4 Mace, G.M., 2010. Distorted views of biodiversity: spatial and temporal bias in species
5 occurrence data. *PLoS biology*, 8(6), p.e1000385.
6
7
8 +Bradie, J. and Leung, B., 2017. A quantitative synthesis of the importance of variables used in
9 MaxEnt species distribution models. *Journal of Biogeography*, 44(6), pp.1344-1361.
10
11 +Brotons, L., Thuiller, W., Araújo, M.B. and Hirzel, A.H., 2004. Presence- absence versus
12 presence- only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4),
13 pp.437-448.
14
15
16 +Buisson, L., Thuiller, W., Casajus, N., Lek, S. and Grenouillet, G., 2010. Uncertainty in
17 ensemble forecasting of species distribution. *Global Change Biology*, 16(4), pp.1145-1157.
18
19
20 +Carrasco, L., Giam, X., Papeş, M. and Sheldon, K.S., 2019. Metrics of Lidar-Derived 3D
21 Vegetation Structure Reveal Contrasting Effects of Horizontal and Vertical Forest Heterogeneity
22 on Bird Species Richness. *Remote Sensing*, 11(7), p.743.
23
24
25 + Cody, M.L. ed., 1985. *Habitat selection in birds*. Academic Press.
26
27
28 +Coops, N.C., Rickbeil, G.J., Bolton, D.K., Andrew, M.E. and Brouwers, N.C., 2018.
29 Disentangling vegetation and climate as drivers of Australian vertebrate richness. *Ecography*,
30 41(7), pp.1147-1160.
31
32
33 +Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., Millán, V.E.G., Strasser, T.,
34 Borre, J.V., Toon, S. and Michael, F., 2015. Remote sensing for mapping natural habitats and
35 their conservation status—New opportunities and challenges. *International Journal of Applied*
36 *Earth Observation and Geoinformation*, 37, pp.7-16.
37
38
39 +Cortez, P. and Embrechts, M.J., 2013. Using sensitivity analysis and visualization techniques
40 to open black box data mining models. *Information Sciences*, 225, pp.1-17.
41
42
43 +Cortez, P., 2016. *rminer: data mining classification and regression methods*. R package
44 version 1.4.2.
45
46
47 +Curry, C.M., Ross, J.D., Contina, A.J. and Bridge, E.S., 2018. Varying dataset resolution alters
48 predictive accuracy of spatially explicit ensemble models for avian species distribution. *Ecology*
49 *and evolution*, 8(24), pp.12867-12878.
50
51
52 +Davies, A.B. and Asner, G.P., 2014. Advances in animal ecology from 3D-LiDAR ecosystem
53 mapping. *Trends in ecology & evolution*, 29(12), pp.681-691.
54
55
56 +Duan, R.Y., Kong, X.Q., Huang, M.Y., Fan, W.Y. and Wang, Z.G., 2014. The predictive
57 performance and stability of six species distribution models. *PLoS One*, 9(11), p.e112764.
58
59
60

- 1
2
3
4 +Dubayah, R. O., Blair, J. B., Goetz, S. J., Fatoyinbo, L., Hansen, M. C., Healey, S. P., Hofton,
5 M., Hurtt, G., Kellner, J. R., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S.,
6 Jantz, P., Marselis, S., Patterson, P., Qi, W., Silva, C. (2020). The Global Ecosystem Dynamics
7 Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of*
8 *Remote Sensing* (in press).
- 9
10 +Engler, J.O., Stiels, D., Schidelko, K., Strubbe, D., Quillfeldt, P. and Brambilla, M., 2017. Avian
11 SDMs: current state, challenges, and opportunities. *Journal of avian biology*, 48(12), pp.1483-
12 1504.
- 13
14
15 +Fick, S.E. and Hijmans, R.J., 2017. WorldClim 2: new 1- km spatial resolution climate surfaces
16 for global land areas. *International journal of climatology*, 37(12), pp.4302-4315.
- 17
18
19 +Fielding, A.H. and Bell, J.F., 1997. A review of methods for the assessment of prediction errors
20 in conservation presence/absence models. *Environmental conservation*, 24(1), pp.38-49.
- 21
22
23 +Flint, L.E., Flint, A.L., Thorne, J.H. and Boynton, R., 2013. Fine-scale hydrologic modeling for
24 regional landscape applications: the California Basin Characterization Model development and
25 performance. *Ecological Processes*, 2(1), p.25.
- 26
27
28 +Gavish, Y., Marsh, C.J., Kuemmerlen, M., Stoll, S., Haase, P. and Kunin, W.E., 2017.
29 Accounting for biotic interactions through alpha- diversity constraints in stacked species
30 distribution models. *Methods in Ecology and Evolution*, 8(9), pp.1092-1102.
- 31
32
33 +Gastón, A., Ciudad, C., Mateo-Sanchez, M.C., Garcia-Vinas, J.I., López-Leiva, C., Fernández-
34 Landa, A., Marchamalo, M., Cuevas, J., de la Fuente, B., Fortin, M.J. and Saura, S., 2017.
35 Species' habitat use inferred from environmental variables at multiple scales: how much we gain
36 from high-resolution vegetation data?. *International Journal of Applied Earth Observation and*
37 *Geoinformation*, 55, pp.1-8.
- 38
39
40 +Goetz, S.J., Steinberg, D., Dubayah, R. and Blair, B., 2007. Laser remote sensing of canopy
41 habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest,
42 USA. *Remote Sensing of Environment*, 108(3), pp.254-263.
- 43
44
45 +Goetz, S.J., Steinberg, D., Betts, M.G., Holmes, R.T., Doran, P.J., Dubayah, R. and Hofton,
46 M., 2010. Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird.
47 *Ecology*, 91(6), pp.1569-1576.
- 48
49
50 +Goetz, S. J., Sun, M., Zolkos, S., Hansen, A., & Dubayah, R. 2014. The relative importance of
51 climate and vegetation properties on patterns of North American breeding bird species richness.
52 *Environmental Research Letters*, 9(3), 034013.
- 53
54
55
56
57
58
59
60

- 1
2
3 +Hao, T., Elith, J., Guillera- Arroita, G. and Lahoz- Monfort, J.J., 2019. A review of evidence
4 about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity*
5 and *Distributions*.
6
7
8 +Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R. and
9 Dubayah, R., 2019. The GEDI Simulator: A Large- Footprint Waveform Lidar Simulator for
10 Calibration and Validation of Spaceborne Missions. *Earth and Space Science*, 6(2), pp.294-310.
11
12
13 +Hawkins, B.A., Porter, E.E. and Felizola Diniz-Filho, J.A., 2003. Productivity and history as
14 predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology*, 84(6), pp.1608-1623.
15
16
17 +He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.N., Schmidtlein, S., Turner, W.,
18 Wegmann, M. and Pettorelli, N., 2015. Will remote sensing shape the next generation of
19 species distribution models?. *Remote Sensing in Ecology and Conservation*, 1(1), pp.4-18.
20
21
22 + Hill, R. A., S. A. Hinsley, D. L. A. Gaveau, and P. E. Bellamy. 2004. Predicting habitat quality
23 for great tits. *International Journal of Remote Sensing* 25:4851–4855.
24
25
26 +Hobi, M.L., Dubinin, M., Graham, C.H., Coops, N.C., Clayton, M.K., Pidgeon, A.M. and
27 Radeloff, V.C., 2017. A comparison of Dynamic Habitat Indices derived from different MODIS
28 products as predictors of avian species richness. *Remote sensing of environment*, 195, pp.142-
29 152.
30
31
32 +Hof, A.R., Jansson, R. and Nilsson, C., 2012. The usefulness of elevation as a predictor
33 variable in species distribution modelling. *Ecological Modelling*, 246, pp.86-90.
34
35
36 +Hofton, M.A., Minster, J.B. and Blair, J.B., 2000. Decomposition of laser altimeter waveforms.
37 *IEEE Transactions on geoscience and remote sensing*, 38(4), pp.1989-1996.
38
39
40 +Howard, C., Stephens, P.A., Pearce- Higgins, J.W., Gregory, R.D. and Willis, S.G., 2015. The
41 drivers of avian abundance: patterns in the relative importance of climate and land use. *Global*
42 *Ecology and Biogeography*, 24(11), pp.1249-1260.
43
44
45 +Huang, Q., Swatantran, A., Dubayah, R. and Goetz, S.J., 2014. The influence of vegetation
46 height heterogeneity on forest and woodland bird species richness across the United States.
47 *PLoS One*, 9(8), p.e103236.
48
49
50 +Hunter, M.L. and Hunter Jr, M.L. eds., 1999. *Maintaining biodiversity in forest ecosystems*.
51 Cambridge university press.
52
53
54 +IPBES, 2019. Global assessment report on biodiversity and ecosystem services of the
55 Intergovernmental Science- Policy Platform on Biodiversity and Ecosystem Services. E. S.
56 Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES Secretariat, Bonn, Germany.
57
58
59
60

- 1
2
3 +Jackson, H.B. and Fahrig, L., 2015. Are ecologists conducting research at the optimal scale?.
4 *Global Ecology and Biogeography*, 24(1), pp.52-63.
5
6
7 + Jensen, J.R., 1996. Introductory digital image processing: a remote sensing perspective (No.
8 Ed. 2). Prentice-Hall Inc.
9
10 +Jetz, W., Wilcove, D.S. and Dobson, A.P., 2007. Projected impacts of climate and land-use
11 change on the global diversity of birds. *PLoS biology*, 5(6).
12
13
14 +Kadmon, R., Farber, O. and Danin, A., 2003. A systematic analysis of factors affecting the
15 performance of climatic envelope models. *Ecological Applications*, 13(3), pp.853-867.
16
17
18 + Kearney, M.R., Gillingham, P.K., Bramer, I., Duffy, J.P. and Maclean, I.M., 2019. A method for
19 computing hourly, historical, terrain-corrected microclimate anywhere on Earth. *Methods in*
20 *Ecology and Evolution*.
21
22
23 +Lawler, J.J., Shafer, S.L., White, D., Kareiva, P., Maurer, E.P., Blaustein, A.R. and Bartlein,
24 P.J., 2009. Projected climate- induced faunal change in the Western Hemisphere. *Ecology*,
25 90(3), pp.588-597.
26
27
28 + Lembrechts, J.J., Nijs, I. and Lenoir, J., 2019. Incorporating microclimate into species
29 distribution models. *Ecography*, 42(7), pp.1267-1279.
30
31
32 + Levin, S.A., 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur
33 award lecture. *Ecology*, 73(6), pp.1943-1967.
34
35 +MacArthur, R.H. and MacArthur, J.W., 1961. On bird species diversity. *Ecology*, 42(3), pp.594-
36 598.
37
38
39 +Manzoor, S.A., Griffiths, G. and Lukac, M., 2018. Species distribution model transferability and
40 model grain size—finer may not always be better. *Scientific reports*, 8(1), p.7168.
41
42
43 +Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. and Thuiller, W., 2009. Evaluation of
44 consensus methods in predictive species distribution modelling. *Diversity and distributions*,
45 15(1), pp.59-69.
46
47
48 +Martínez- Freiria, F., Tarroso, P., Rebelo, H. and Brito, J.C., 2016. Contemporary niche
49 contraction affects climate change predictions for elephants and giraffes. *Diversity and*
50 *Distributions*, 22(4), pp.432-444.
51
52 + Mayor, S.J., Schneider, D.C., Schaefer, J.A. and Mahoney, S.P., 2009. Habitat selection at
53 multiple scales. *Ecoscience*, 16(2), pp.238-247.
54
55
56
57
58
59
60

- 1
2
3 +McClure, C.J., Korte, A.C., Heath, J.A. and Barber, J.R., 2015. Pavement and riparian forest
4 shape the bird community along an urban river corridor. *Global Ecology and Conservation*, 4,
5 pp.291-310.
6
7
8 +McPherson, J. and Jetz, W., 2007. Effects of species' ecology on the accuracy of distribution
9 models. *Ecography*, 30(1), pp.135-151.
10
11
12 +Mertes, K. and Jetz, W., 2018. Disentangling scale dependencies in species environmental
13 niches and distributions. *Ecography*, 41(10), pp.1604-1615.
14
15
16 +Mod, H.K., Scherrer, D., Luoto, M. and Guisan, A., 2016. What we use is not what we know:
17 environmental predictors in plant distribution models. *Journal of Vegetation Science*, 27(6),
18 pp.1308-1322.
19
20 Müller, J., Stadler, J., and Brandl, R. (2010). Composition versus physiognomy of vegetation as
21 predictors of bird assemblages: The role of lidar. *Remote Sensing of Environment*, 114(3), pp.
22 490-495.
23
24
25 +National Academies of Sciences, Engineering, and Medicine (NASEM), 2018. Thriving on Our
26 Changing Planet: A Decadal Strategy for Earth Observation from Space. The National
27 Academies Press, Washington, DC. <http://dx.doi.org/10.17226/24938>.
28
29
30 +Neuenschwander, A. and Pitts, K., 2019. The ATL08 land and vegetation product for the
31 ICESat-2 Mission. *Remote sensing of environment*, 221, pp.247-259.
32
33
34 +Pearce, J. and Ferrier, S., 2000. Evaluating the predictive performance of habitat models
35 developed using logistic regression. *Ecological modelling*, 133(3), pp.225-245.
36
37
38 +Pearson, R.G. and Dawson, T.P., 2003. Predicting the impacts of climate change on the
39 distribution of species: are bioclimate envelope models useful?. *Global ecology and*
40 *biogeography*, 12(5), pp.361-371.
41
42
43 +Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford,
44 M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C. and Coops, N.C., 2013. Essential
45 biodiversity variables. *Science*, 339(6117), pp.277-278.
46
47
48 +Popescu, S.C., Zhou, T., Nelson, R., Neuenschwander, A., Sheridan, R., Narine, L. and
49 Walsh, K.M., 2018. Photon counting LiDAR: An adaptive ground and canopy height retrieval
50 algorithm for ICESat-2 data. *Remote sensing of environment*, 208, pp.154-170.
51
52
53 +Proença, V., Martin, L.J., Pereira, H.M., Fernandez, M., McRae, L., Belnap, J., Böhm, M.,
54 Brummitt, N., García-Moreno, J., Gregory, R.D. and Honrado, J.P., 2017. Global biodiversity
55 monitoring: from data sources to essential biodiversity variables. *Biological Conservation*, 213,
56 pp.256-263.
57
58
59
60

1
2
3
4 +R Core Team (2019). R: A language and environment for statistical computing. R Foundation
5 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

6
7
8 +Radeloff, V.C., Dubinin, M., Coops, N.C., Allen, A.M., Brooks, T.M., Clayton, M.K., Costa,
9 G.C., Graham, C.H., Helmers, D.P., Ives, A.R. and Kolesov, D., 2019. The Dynamic Habitat
10 Indices (DHIs) from MODIS and global biodiversity. *Remote Sensing of Environment*, 222,
11 pp.204-214.

12
13
14 +Rechsteiner, C., Zellweger, F., Gerber, A., Breiner, F.T. and Bollmann, K., 2017. Remotely
15 sensed forest habitat structures improve regional species conservation. *Remote Sensing in
16 Ecology and Conservation*, 3(4), pp.247-258.

17
18
19 +Rocchini, D., Andreo, V., Förster, M., Garzon-Lopez, C.X., Gutierrez, A.P., Gillespie, T.W.,
20 Hauffe, H.C., He, K.S., Kleinschmit, B., Mairota, P. and Marcantonio, M., 2015. Potential of
21 remote sensing to predict species invasions: A modelling perspective. *Progress in Physical
22 Geography*, 39(3), pp.283-309.

23
24
25 +Rodewald, P. (Editor). 2015. *The Birds of North America*: <https://birdsna.org>. Cornell
26 Laboratory of Ornithology, Ithaca, NY.

27
28
29 +Rosenberg, K.V., Dokter, A.M., Blancher, P.J., Sauer, J.R., Smith, A.C., Smith, P.A., Stanton,
30 J.C., Panjabi, A., Helft, L., Parr, M. and Marra, P.P., 2019. Decline of the North American
31 avifauna. *Science*, 366(6461), pp.120-124.

32
33
34 +Rotenberry, J.T., 1985. The role of habitat in avian community composition: physiognomy or
35 floristics?. *Oecologia*, 67(2), pp.213-217.

36
37
38 +Rottenborn, S.C., 1999. Predicting the impacts of urbanization on riparian bird communities.
39 *Biological conservation*, 88(3), pp.289-299.

40
41 +Sauer, J. R., D. K. Niven, J. E. Hines, D. J. Ziolkowski, Jr, K. L. Pardieck, J. E. Fallon, and W.
42 A. Link. 2017. *The North American Breeding Bird Survey, Results and Analysis 1966 - 2015*.
43 Version 2.07.2017 USGS Patuxent Wildlife Research Center, Laurel, MD.

44
45
46 +Seavy, N.E., Viers, J.H. and Wood, J.K., 2009. Riparian bird response to vegetation structure:
47 a multiscale analysis using LiDAR measurements of canopy height. *Ecological Applications*,
48 19(7), pp.1848-1857.

49
50
51 +Sohl, T.L., 2014. The relative impacts of climate and land-use change on conterminous United
52 States bird species from 2001 to 2075. *PLoS One*, 9(11), p.e112251.

- 1
2
3 +Stevens, B.S. and Conway, C.J., 2019. Predicting species distributions: unifying model
4 selection and scale optimization for multi- scale occupancy models. *Ecosphere*, 10(5),
5 p.e02748.
6
7
8 +Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., Alexander, J.D., Wiens, J.A. and
9 Root, T.L., 2009. Re-shuffling of species with climate disruption: a no-analog future for
10 California birds?. *PLoS one*, 4(9), p.e6825.
11
12
13 +Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. and Kelling, S., 2009. eBird: A
14 citizen-based bird observation network in the biological sciences. *Biological Conservation*,
15 142(10), pp.2282-2292.
16
17
18 +Swatantran, A., Dubayah, R., Goetz, S., Hofton, M., Betts, M.G., Sun, M., Simard, M. and
19 Holmes, R., 2012. Mapping migratory bird prevalence using remote sensing data fusion. *PLoS*
20 *one*, 7(1), p.e28922.
21
22
23 +Sulla-Menashe, D., Gray, J.M., Abercrombie, S.P. and Friedl, M.A., 2019. Hierarchical
24 mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover
25 product. *Remote sensing of environment*, 222, pp.183-194.
26
27
28 +Syphard, A.D. and Franklin, J., 2010. Species traits affect the performance of species
29 distribution models for plants in southern California. *Journal of Vegetation Science*, 21(1),
30 pp.177-189.
31
32
33 +Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D.B. and Blair, B.,
34 2012. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva,
35 Costa Rica. *Remote Sensing of Environment*, 124, pp.242-250.
36
37
38 +Tattoni, C., Rizzolli, F. and Pedrini, P., 2012. Can LiDAR data improve bird habitat suitability
39 models?. *Ecological Modelling*, 245, pp.103-110.
40
41
42 +Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. and Guisan, A., 2014. Measuring
43 the relative effect of factors affecting species distribution model predictions. *Methods in Ecology*
44 *and Evolution*, 5(9), pp.947-955.
45
46
47 +Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. and Legendre, F., 2017. Taxonomic
48 bias in biodiversity data and societal preferences. *Scientific reports*, 7(1), pp.1-14.
49
50
51 +Tsui, O. W., Coops, N. C., Wulder, M. A., & Marshall, P. L. 2013. Integrating airborne LiDAR
52 and space-borne radar via multivariate kriging to estimate above-ground biomass. *Remote*
53 *Sensing of Environment*, 139, 340-352.
54
55 +U.S. Geological Survey, 2019. Breeding Bird Atlas Explorer (online resource). Accessed 29
56 May 2019. <http://www.pwrc.usgs.gov/bba>.
57
58
59
60

- 1
2
3
4 +Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S. and Clawges, R.M., 2008. Lidar:
5 shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the*
6 *Environment*, 6(2), pp.90-98.
7
8
9 +Wan, H.Y., McGarigal, K., Ganey, J.L., Lauret, V., Timm, B.C. and Cushman, S.A., 2017.
10 Meta-replication reveals nonstationarity in multi-scale habitat selection of Mexican Spotted Owl.
11 *The Condor: Ornithological Applications*, 119(4), pp.641-658.
12
13
14 + Whittaker, R.J., Willis, K.J. and Field, R., 2001. Scale and species richness: towards a
15 general, hierarchical theory of species diversity. *Journal of biogeography*, 28(4), pp.453-470.
16
17
18 +Wiens, J.A., 1989. Spatial scaling in ecology. *Functional ecology*, 3(4), pp.385-397.
19
20 + Wilson, A.M. and Jetz, W., 2016. Remotely sensed high-resolution global cloud dynamics for
21 predicting ecosystem and biodiversity distributions. *PLoS biology*, 14(3), p.e1002415.
22
23
24 + Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D. and Coomes, D., 2019. Advances in
25 microclimate ecology arising from remote sensing. *Trends in ecology & evolution*, 34(4), pp.327-
26 341.
27
28
29 +Zurell, D., Thuiller, W., Pagel, J., Cabral, J.S., Münkemüller, T., Gravel, D., Dullinger, S.,
30 Normand, S., Schiffers, K.H., Moore, K.A. and Zimmermann, N.E., 2016. Benchmarking novel
31 approaches for modelling species range dynamics. *Global change biology*, 22(8), pp.2651-
32 2664.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60