

AN ABSTRACT OF THE THESIS OF

Samuel Hooper for the degree of Master of Science in Geography presented on January 5, 2017.

Title: Spatial and Temporal Dynamics of Broad-scale Predictive Models: Influences of Scale

Abstract approved: _____

Robert E. Kennedy

Developing accurate predictive distribution models requires adequately representing relevant spatial and temporal scales, as these scales are ultimately reflective of the relationships between distributions and influential environmental conditions. In this research, we considered both spatial and temporal scale and the influence each has on predicting broad-scale distributions of two disparate but related phenomena: land cover and bird distributions. Employing machine-learning algorithms, we first developed land cover time series datasets covering all of California, Oregon, and Washington with a model that simultaneously reflects local-scale heterogeneity and broad-scale homogeneity. We then used these and other land cover time series datasets to investigate the effects of temporal resolution on species distribution models.

In the second chapter, we focused on the importance of accurately representing the spatial scale of relationships between predictors and a response variable for broad-scale predictive models. Using both a novel machine-learning algorithm and a novel predictor dataset, we developed dense time series forest canopy cover (FCC) and impervious surface cover (ISC) datasets at a 30-meter spatial resolution for all of California, Oregon, and Washington. To develop both datasets, we employed a spatial ensemble modeling method using a population of locally defined and

spatially overlapping decision trees, making it both appropriate at continental-scales and sensitive to local variation in predictor-response relationships. Our predictor variables were products of LandTrendr, a tool for developing time series images and derivatives from the Landsat archive. To develop the most accurate time series of FCC and ISC, we first tested two model parameters, sample size and estimator size. Using the best-performing configuration of each, we then compared our models with locally defined estimators to bagged decision trees, the most comparable model with globally defined estimators. Using the best-performing models and LandTrendr imagery, we developed yearly FCC and ISC maps, spanning 1990-2012. To test the temporal extensibility of our models, we compared our predicted 2011 maps to 2011 maps from the National Land Cover Database. We found that model performance for both FCC and ISC decreased with increasing estimator size and that models with locally defined estimators outperformed bagged decision trees. We also found that our models performed well when extending learned predictor-response relationships to predict 2011 FCC and ISC distributions. These results, in concert with several novel byproducts of the models that we developed, demonstrate that representing local-scale spatial relationships is critical to producing accurate broad-scale distribution models.

In the third chapter, we investigated the influence of temporal scale on an avian species distribution model (SDM) by comparing models developed with different temporal resolutions of land cover predictor data. We expressed temporal resolution as the time interval between representative dates of predictor data. By adjusting the temporal resolution of land cover predictors but still using yearly response data, models at coarser resolutions were potentially trained with bird observations that did not match the land cover contemporaneous with

observations. We tested different temporal resolutions for five bird species to capture how habitat preference may affect SDM response to temporal resolution. Our results confirmed that temporal resolution has a noticeable effect on SDMs, but in unexpected ways. For three of the five species, the best-performing models were produced at the coarsest temporal resolution, with other species showing better performance at moderate resolutions. We also found only subtle evidence supporting the idea that habitat preference influenced SDM response to temporal resolution. These results indicate that 1) temporal dynamics of avian species-environment relationships are highly complex and specific to individual species and 2) that representation of temporal scale has a prominent effect on model outcome.

©Copyright by Samuel Hooper
January 5, 2017
All Rights Reserved

Spatial and Temporal Dynamics of Broad-scale Predictive Models: Influences of Scale

by
Samuel Hooper

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented January 5, 2017
Commencement June 2017

Master of Science thesis of Samuel Hooper presented on January 5, 2017

APPROVED:

Major Professor, representing Geography

Dean of the College of Earth, Ocean, and Atmospheric Sciences

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Samuel Hooper, Author

ACKNOWLEDGEMENTS

As with all scientific endeavors, this thesis stands as a small contribution in a long line of questioning, which has led to the intellectual and technological legacy that I inherited. In particular, this research relied heavily on the diligent and creative efforts of others who have worked for years to develop LandTrendr and derivative products. I am thankful for Neeti and Tara Larrue and their work in developing the classified land cover time series that I used in Chapter 3. Tara also provided sage programming and computing advice beyond what might be expected from someone so early in her computer science career.

This thesis would also not have been possible without the dedication of so many to open source code and open data. Members of the open source community have volunteered countless hours to develop Python libraries and contribute to the wealth of knowledge accessible with the click of a mouse. This research was also only feasible because of the innumerable hours of bird observations provided by eBird participants and the equally admirable efforts of the eBird team and the Cornell Lab of Ornithology, maintaining this rich and publicly available dataset. This work was also supported in part by a graduate fellowship under USGS Contract G15AC00018, a Landsat Science Team project awarded to PI Robert Kennedy.

Instrumental to this work has of course been the advice, guidance, and expertise of several faculty members and staff at Oregon State, both on my committee and in the broader OSU community. As soon as I started working with Robert Kennedy, I knew I was lucky to have such a great advisor, mainly because almost every person to whom I mentioned his name told me

exactly that. The longer I have worked with him, the more he has confirmed my initial impression. On top of his ability to turn on a dime concerning every statistics, image processing, or ecology question with which I ever approached him (which were many), Robert was always supportive, caring, and incredibly generous with his time. Doug Robinson provided invaluable counsel on bird ecology and creative ways to address the complications of crowd-sourced ecological observations. I was always impressed with and grateful for Denis White's enthusiasm and thoughtful insights that made me consider new directions and lines of questioning throughout the development of this thesis. I am also thankful for the volunteered time and effort of my graduate committee representative, Tiffany Garcia. Julia Jones, while not directly involved with this research, continued to provide support, counsel, and guidance beginning even before I started my graduate studies at OSU through to the culmination of this thesis. Her generosity with her time, a resource in incredibly high demand both within and outside of the Geography program, was impressive. I am also grateful for all of the administrative staff members who have made my graduate education so much easier, including Lori Hartline, Robert Allan, Stacey Schulte, Melinda Jensen, Renee Freeman, and many others with whom I have not interacted with directly.

Graduate school would have also been far more difficult if it were not for all of the friendship, family support, and memorable adventures I have been fortunate enough to have over the last two years. In particular, I am grateful for the encouragement of my loving parents, Murray Hooper, Judy Shub-Condliffe, and Jack Condliffe, throughout all of my various endeavors. To Katie and Christopher Smith, I am glad that we all wound up in the same town, even if for only a few short years. Thank you, as well, to Dave Leydet, Kate Jones, Melva Treviño-Peña, Melissa

McCracken, and Dan Stephen who made up a fun and supportive Geography cohort. I am particularly grateful to have had an adventure buddy like Dave, whose ski touring, mountaineering, and otherwise outdoor enthusiasm was one of the only reasons I was able to maintain some semblance of a work-life balance. I would also like to thank my wonderfully supportive officemates, Jane Darbyshire, Dan Stephen (again), and Tyler Harris, for providing a welcoming and creative atmosphere for sharing programming and visualization ideas. Thank you, as well, to so many others for inviting and joining me on all of the other much-needed adventures and socializing.

Lastly, many people successfully make it through graduate school without the loving companionship I have been fortunate enough to share with my partner, Jen Johnston, but I am not sure how. We knew graduate school would be an adventure of a sort to which we were unaccustomed, and I am glad to have shared the experience with you.

TABLE OF CONTENTS

	<u>Page</u>
Chapter 1: Introduction	1
References	5
Chapter 2: Global Patterns from Local Signals: Novel Methods for Predicting a Broad-scale Land Cover Time-series.....	9
Introduction	9
Methods	12
Study area	12
Overview of methods	13
Training data	14
Predictor data	15
Model description.....	16
Evaluating map agreement	18
Sample size.....	20
Support set size	20
Comparison of a STEM to a conventional bagged decision tree model	21
Temporal extensibility	21
Spatial assessment of error, uncertainty, and variable importance	21
Results	23
Overall model performance.....	23
Sample size.....	25
Support set size	25
Comparison of a STEM to a BDT.....	26

TABLE OF CONTENTS (continued)

	<u>Page</u>
Temporal extensibility	27
Spatial assessment of error, uncertainty, and variable importance	28
Discussion.....	29
Sample size.....	30
Size of support sets.....	32
STEM vs. BDT.....	34
Temporal extensibility	34
Spatial assessment of error, uncertainty, and variable importance	36
Conclusion	40
References	43
Figures	50
Tables	60
Chapter 3: Influences of Temporal Resolution on Avian Species Distribution Models.....	63
Introduction	63
Methods	67
Study area.....	67
Bird observation data	68
Predictor data	70
Model description.....	72
Adjusting temporal resolution	74
Model evaluation.....	75

TABLE OF CONTENTS (continued)

	<u>Page</u>
Results	78
Discussion.....	81
Potential sources of error	82
SDM response to temporal resolution	83
SDM response across habitat preferences	86
Model performance across different evaluation metrics	89
Conclusion	91
References	93
Figures	100
Tables	108
Chapter 4: Conclusion.....	109
References	119
References	122
Appendix A: Confusion Matrices for STEM Models with Medium and Large Support Sets....	134

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1. Examples of small, medium, and large support sets	50
2.2. Difference images between STEM and NLCD maps in San Francisco, CA and Seattle, WA.	51
2.3. Illustration of computing an OOB map from a STEM.	52
2.4. Two-dimensional histograms comparing STEM predictions of FCC and ISC to NLCD reference samples using the best-match method and average method.	53
2.5. All five summary statistics for the FCC and ISC BDTs and the best performing STEMs....	54
2.6. Difference of STEM and BDT maps averaged by 35,000 ha hexagons	55
2.7. Standard deviation and OOB score maps for FCC and ISC models.....	56
Figure 2.8. Variable importance maps for FCC and ISC.....	57
2.9. Cumulative distribution functions for 2011 FCC STEM mode, STEM mean, and NLCD...	58
2.10. Error in source data: FCC NLCD cross-validation error, ISC NLCD cross-validation error, and 2001 LandTrendr mean squared error.....	59
3.1. Land cover change by area between 2001 and 2011 from the National Land Cover Database	100
3.2. Density of ERD observations within a 1 km radius.....	101
3.3. Model performance according to the OOB score, AUC, and RMSE for four different temporal resolutions.....	102
3.4. RMSE and AUC by land cover type for four different temporal resolutions for all five species.	103
3.5. Occurrence probabilities for all five species averaged within 35,000 ha hexagons	104
3.6. Mean relative importance for all species of constant predictors and yearly land cover predictors for each temporal resolution	105
3.7. Temporal distribution of June observations in the ERD.....	106

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
3.8. Difference of standard deviations between 1-year and 10-year Black-headed Grosbeak model.....	107

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1. Area-estimated confusion matrix for the best performing FCC model	60
2.2. Area-estimated confusion matrix for the best performing ISC model.....	61
2.3. Summary statistics for all FCC and ISC models	62
3.1. Predictors used for SDMs	108

LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
A.1. Confusion matrix for medium-sized (400 km x 300 km) forest canopy cover model.....	134
A.2. Confusion matrix for large-sized (800 km x 600 km) forest canopy cover model.....	135
A.3. Confusion matrix for medium-sized (400 km x 300 km) impervious surface cover model	136
A.4. Confusion matrix for large-sized (800 km x 600 km) impervious surface cover model	137

CHAPTER 1: INTRODUCTION

Understanding the spatial and temporal distributions of environmental phenomena is a common challenge in ecology. Accurately predicting distributions requires developing a model that can decipher the relationships between influential conditions and these distributions. To adequately estimate such predictor-response relationships, proper representation of the spatial and temporal scales important to a phenomenon is necessary; the scale at which a model is developed can have prominent effects on the outcome of the model (Austin and Van Neil 2011; Gottschalk et al. 2011; Song et al. 2013). For many applications, differences in the scale of a model are often expressed as the resolution, or grain size, of some element of the model. Many authors have argued that resolution must match scales important to the phenomena considered (Addicott et al. 1987; Betts et al. 2006; Mitchel et al. 2001).

Increasingly, managers and conservation strategists must address problems at continental and global scales. Adequately meeting this challenge requires data processed in a consistent manner, making broad-scale data particularly valuable (Gray and Song 2013; White et al. 2014; Wang et al. 2015). Furthermore, as the extent of a modeling region under consideration increases, the relationships between predictors and a response variable are likely to vary across the entire region (Fink et al. 2010). For instance, individuals of a local subpopulation of a given species may exploit certain local resources, but others in a different part of its range may occupy a different niche (Murphy and Lovett-Doust 2007; Peterson and Holt 2003; Quevedo, Svanba, and Eklo 2009), indicating that some predictor-response relationships are scale-specific. Practically, the spatial scales important to a particular phenomenon can be understood as the distance beyond

which the relationships between that phenomenon and the environmental conditions that affect it have changed. Modeling and analysis methods must, therefore, account for such local-scale effects to avoid spuriously applied relationships (Fink et al. 2010; Ashcroft et al. 2011). Most spatial distribution models, however, ignore local-scale patterns and, instead, apply the same predictor-response relationships across an entire modeling region.

In the temporal domain, the question of an appropriate scale is similar to spatial scale: how much time can pass (i.e., temporal distance) before the state of a phenomenon is no longer related to some set of original environmental conditions? Spatial patterns are often indelibly tied to temporal patterns (Adler and Lauenroth 2003; Ives and Klopfer 1997), so consideration of temporal scales is of equal importance as spatial scales. This is particularly true with models that describe distributions of phenomena that are highly temporally dynamic or that may respond to temporally dynamic predictors, such as land cover (see Homer et al. 2015). Ecologists have long recognized the importance of spatial resolution on species-environment analyses, but the influence of temporal resolution remains generally unexplored (Elith and Leathwick 2009, 689; Radchuk et al. 2014). The logistical limitations of field data collection and the unavailability of dense time series data, however, often require that researchers develop spatial distribution models using predictor data with temporal resolutions or extents that do not match response data (e.g., Fink et al. 2010; Gschweng et al. 2012; Gutzwiller and Barrow 2001; Rullman and Marzluff 2014). The potential effects of coarse or mismatched temporal resolutions for predictive models are largely unexamined.

Both spatial and temporal scales are therefore of fundamental importance when developing predictive models of distributions, but there are many questions concerning how best to express them across broad spatial and temporal extents. This thesis explored the dynamics of predictive models covering broad extents by considering scale in both the spatial and temporal domains. In Chapter 2, we focused on the importance of accurately representing the spatial scale of relationships between predictors and a response variable for broad-scale predictive models. We applied a machine-learning ensemble algorithm in a novel context and with a novel dataset to develop land cover models. We ultimately used those models to develop time series land cover maps, spanning more than two decades. Chapter 3 was an investigation of temporal scale and its influence on species distribution models. We used the land cover time series developed in Chapter 2 and other land cover data to examine how the temporal dynamics of land cover affect model performance. Together, both chapters represent an effort to develop more accurate representations of spatiotemporal distributions.

In Chapter 2, we addressed the problem of variable predictor-response relationships at broad scales by using a spatial ensemble model comprised of a population of locally defined, spatially overlapping estimators (i.e., base models). By constraining the spatial extent of each estimator, each one makes a prediction using predictor-response relationships from a limited area, producing a broad-scale model that remains sensitive to local variation in those relationships. The optimal estimator size is a reflection of the spatial scale important to the phenomenon under consideration. The predictor data we used in this chapter were mostly derived from LandTrendr, an approach for developing time series images and land cover change maps from the Landsat archive (Kennedy et al. 2012). Leveraging this novel modeling method and spectral time series

dataset, we developed yearly time series of two influential land cover metrics, forest canopy cover (FCC) and impervious surface cover (ISC), for all of California, Oregon, and Washington at a 30 m spatial resolution. Spanning from 1990 to 2012, yearly datasets depicting these phenomena at such fine spatial resolution and broad spatial extent did not previously exist. These data will ultimately aid managers and researchers in understanding how these influential metrics of land cover patterns affect a wide range of ecological, geochemical, and social processes.

Chapter 3 explored the influence of differing temporal resolutions between land cover predictors and response data on a species distribution model (SDM). Land cover is often central to the development of SDMs (Manel et al. 2000; Mitchel et al. 2001; Rullman and Marzluff 2014). Land cover is also temporally dynamic, so accurately representing changes in land cover at an appropriate temporal scale is critical for an SDM. We again used LandTrendr data and another novel dataset, eBird, to examine the effects of temporal resolution on an SDM. eBird is a citizen science project comprising an online database of bird observation checklists submitted by users worldwide (Dickinson et al. 2010). This dataset is unique among bird observation databases in both its fine resolution and broad extent—both spatial and temporal (Hochachka and Fink 2012). In this chapter, we applied a random forest, another machine learning ensemble algorithm, to model breeding season distributions of five different bird species. Although a random forest uses global estimators that apply predictor-response relationships across the entire modeling region, Fink and others (2010) found that for representing bird distributions when they are more stable (e.g., the breeding season), a model with globally defined estimators performed comparably to a locally sensitive model. Using the random forest algorithm, we developed multiple SDMs with differing temporal resolutions of predictor data. By adjusting the time interval between predictor

datasets but still training each random forest on yearly bird response data, we evaluated how the disparity between the two can affect model performance. We modeled distributions for multiple species to determine if habitat preference was related to performance with differing temporal resolutions, as different land cover types exhibit varied rates of disturbance (Homer et al. 2015). This work will hopefully shed new light on species-environment relationships, and provide insight on appropriate temporal resolutions for SDMs and bird monitoring programs.

Modeling distributions at broad spatial and temporal extents is necessary for effective conservation and management of a variety of resources, but developing such expansive datasets presents unique challenges. This research attempted to resolve two issues concerning appropriate representations of scale for broad-scale models. Spatial and temporal resolutions are demonstrably significant model considerations because, ultimately, they are a reflection of ecological reality. This thesis demonstrated that the development of any distribution model must offer considerable attention to both of these influential parameters.

References

- Addicott, John F., John M. Aho, Michael F. Antolin, Dianna K. Padilla, John S. Richardson, and Daniel A. Soluk. 1987. "Ecological Neighborhoods: Scaling Environmental Patterns." *Oikos* 49 (3): 340–46. doi:10.2307/3565770.
- Adler, Peter B., and William K. Lauenroth. 2003. "The Power of Time: Spatiotemporal Scaling of Species Diversity." *Ecology Letters* 6 (8): 749–56. doi:10.1046/j.1461-0248.2003.00497.x.
- Ashcroft, Michael B., Kristine O. French, and Laurie A. Chisholm. 2011. "An Evaluation of Environmental Factors Affecting Species Distributions." *Ecological Modelling* 222 (3): 524–31. doi:10.1016/j.ecolmodel.2010.10.003.

- Austin, Mike P., and Kimberly P. Van Niel. 2011. "Improving Species Distribution Models for Climate Change Studies: Variable Selection and Scale." *Journal of Biogeography* 38 (1): 1–8. doi:10.1111/j.1365-2699.2010.02416.x.
- Betts, Matthew G., A.W. Diamond, G.J. Forbes, M.-A. Villard, and J.S. Gunn. 2006. "The Importance of Spatial Autocorrelation, Extent and Resolution in Predicting Forest Bird Occurrence." *Ecological Modelling* 191 (2): 197–224. doi:10.1016/j.ecolmodel.2005.04.027.
- Dickinson, Janis L., Benjamin Zuckerberg, and David N. Bonter. 2010. "Citizen Science as an Ecological Research Tool: Challenges and Benefits." *Annual Review of Ecology, Evolution, and Systematics* 41 (1): 149–72. doi:10.1146/annurev-ecolsys-102209-144636.
- Elith, Jane, and John R. Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 677–97. doi:10.1146/annurev.ecolsys.110308.120159.
- Fink, Daniel, Wesley M. Hochachka, Benjamin Zuckerberg, David W. Winkler, Ben Shaby, M. Arthur Munson, Giles Hooker, Mirek Riedewald, Daniel Sheldon, and Steve Kelling. 2010. "Spatiotemporal Exploratory Models for Broad-Scale Survey Data." *Ecological Applications* 20 (8): 2131–2147.
- Franklin, Janet, and Jennifer A Miller. 2009. *Mapping Species Distributions Spatial Inference and Prediction*. Cambridge; New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511810602>.
- Gottschalk, Thomas K., Birgit Aue, Stefan Hotes, and Klemens Ekschmitt. 2011. "Influence of Grain Size on Species–habitat Models." *Ecological Modelling* 222 (18): 3403–12. doi:10.1016/j.ecolmodel.2011.07.008.
- Gray, Josh, and Conghe Song. 2013. "Consistent Classification of Image Time Series with Automatic Adaptive Signature Generalization." *Remote Sensing of Environment* 134 (July): 333–41. doi:10.1016/j.rse.2013.03.022.
- Gschweng, Marion, Elisabeth K. V. Kalko, Peter Berthold, Wolfgang Fiedler, and Jakob Fahr. 2012. "Multi-Temporal Distribution Modelling with Satellite Tracking Data: Predicting Responses of a Long-Distance Migrant to Changing Environmental Conditions." *Journal of Applied Ecology* 49 (4): 803–13. doi:10.1111/j.1365-2664.2012.02170.x.
- Hochachka, Wesley, and Daniel Fink. 2012. "Broad-Scale Citizen Science Data from Checklists: Prospects and Challenges for Macroecology." *Frontiers of Biogeography* 4 (4). <http://escholarship.org/uc/item/01t5c00w>.
- Homer, Collin, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Xian, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. 2015. "Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing

- a Decade of Land Cover Change Information.” *Photogrammetric Engineering & Remote Sensing* 81 (5): 345–54. doi:10.1016/S0099-1112(15)30100-2.
- Ives, Anthony R., and Eric D. Klopfer. 1997. “Spatial Variation in Abundance Created by Stochastic Temporal Variation.” *Ecology* 78 (6): 1907–1913.
- Kennedy, Robert E., Zhiqiang Yang, Warren B. Cohen, Eric Pfaff, Justin Braaten, and Peder Nelson. 2012. “Spatial and Temporal Patterns of Forest Disturbance and Regrowth within the Area of the Northwest Forest Plan.” *Remote Sensing of Environment* 122 (July): 117–33. doi:10.1016/j.rse.2011.09.024.
- Manel, S., S.t. Buckton, and S.j. Ormerod. 2000. “Testing Large-Scale Hypotheses Using Surveys: The Effects of Land Use on the Habitats, Invertebrates and Birds of Himalayan Rivers.” *Journal of Applied Ecology* 37 (5): 756–70. doi:10.1046/j.1365-2664.2000.00537.x.
- Mitchell, Michael S., Richard A. Lancia, and John A. Gerwin. 2001. “Using Landscape-Level Data to Predict the Distribution of Birds on a Managed Forest: Effects of Scale.” *Ecological Applications* 11 (6): 1692–1708. doi:10.1890/1051-0761(2001)011[1692:ULLDTP]2.0.CO;2.
- Murphy, Helen T., and Jon Lovett-Doust. 2007. “Accounting for Regional Niche Variation in Habitat Suitability Models.” *Oikos* 116 (1): 99–110. doi:10.1111/j.2006.0030-1299.15050.x.
- Peterson, A. Townsend, and Robert D. Holt. 2003. “Niche Differentiation in Mexican Birds: Using Point Occurrences to Detect Ecological Innovation.” *Ecology Letters* 6 (8): 774–82. doi:10.1046/j.1461-0248.2003.00502.x.
- Quevedo, Mario, Richard Svanbäck, and Peter Eklöv. 2009. “Intrapopulation Niche Partitioning in a Generalist Predator Limits Food Web Connectivity.” *Ecology* 90 (8): 2263–2274.
- Radchuk, Viktoriia, Karin Johst, Jürgen Groeneveld, Camille Turlure, Volker Grimm, and Nicolas Schtickzelle. 2014. “Appropriate Resolution in Time and Model Structure for Population Viability Analysis: Insights from a Butterfly Metapopulation.” *Biological Conservation* 169 (January): 345–54. doi:10.1016/j.biocon.2013.12.004.
- Rullman, Stan, and John M. Marzluff. 2014. “Raptor Presence Along an Urban–Wildland Gradient: Influences of Prey Abundance and Land Cover.” *Journal of Raptor Research* 48 (3): 257–72. doi:10.3356/JRR-13-32.1.
- Song, Wonkyong, Eunyoung Kim, Dongkun Lee, Mounjin Lee, and Seong-Woo Jeon. 2013. “The Sensitivity of Species Distribution Modeling to Scale Differences.” *Ecological Modelling* 248 (January): 113–18. doi:10.1016/j.ecolmodel.2012.09.012.
- Wang, Jie, Yuanyuan Zhao, Congcong Li, Le Yu, Desheng Liu, and Peng Gong. 2015. “Mapping Global Land Cover in 2001 and 2010 with Spatial-Temporal Consistency at 250m

Resolution.” *ISPRS Journal of Photogrammetry and Remote Sensing* 103 (May): 38–47.
doi:10.1016/j.isprsjprs.2014.03.007.

White, J. C., M. A. Wulder, G. W. Hobart, J. E. Luther, T. Hermosilla, P. Griffiths, N. C. Coops, et al. 2014. “Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science.” *Canadian Journal of Remote Sensing* 40 (3): 192–212.
doi:10.1080/07038992.2014.945827.

CHAPTER 2: GLOBAL PATTERNS FROM LOCAL SIGNALS: NOVEL METHODS FOR PREDICTING A BROAD-SCALE LAND COVER TIME-SERIES

Introduction

Forest canopy cover (FCC) and impervious surface cover (ISC) are two continuous metrics of land cover that are essential to understanding numerous socio-ecological processes. Changes in forest cover are attributed to climatic variation at both regional (Bonan 1997) and global scales (Nemani and Running 1995), and accurate forest cover estimates that are both spatially and temporally explicit are necessary for understanding such processes (Defries et al. 2000).

Martinuzzi and others (2009) demonstrated that accurate estimations of FCC improve species distribution predictions. ISC is an easily understood metric of development intensity (Arnold and Gibbons 1996), yielding a strong influence on hydrologic dynamics including flood risk (Suriya and Mudgal 2012; Campana and Tucci 2001; Sheeder, Ross, and Carlson 2002). ISC also has a demonstrable influence on microclimatic variation from the heat island effect, increasing temperature and reducing evapotranspiration (Carlson and Arthur 2000).

FCC and ISC, then, each exert significant influence on ecosystem function, and data that adequately represent them at fine spatial resolutions can be especially informative. Both metrics, however, are subject to change over time, making fine temporal resolution just as critical to accurate analyses as fine spatial resolution. Assessing changes in land cover requires data representing multiple time periods (Jepson 2005), particularly because the times at which changes may occur is usually unknown. Temporally sparse data, therefore, will not likely represent changes in land cover accurately (Cohen and Goward 2004; Masek et al. 2008).

Several studies have demonstrated the utility of data that are both spatially and temporally detailed (e.g. Goward et al. 2008; Turner et al. 2007). Lunetta and others (2004) recommend a temporal resolution of 1–2 years to adequately assess land cover change.

Predicting spatial distributions for any phenomenon requires discerning relationships between predictor and response variables. In the absence of a priori knowledge of such relationships, spatial distributions are generally modeled using a supervised classification algorithm that, in the process of constructing the model, estimates predictor-response relationships from patterns in training samples. For most models, these learned relationships are then used to interpolate for all other locations, with the same predictor-response relationships applied globally across the entire modeling region.

Models developed at both fine resolutions and over large extents, however, present a unique modeling challenge: variability in relationships between predictor and response variables is likely to increase along with the extent of a modeling region, potentially making learned relationships from one area of the modeling region invalid for another. Increasingly, researchers and managers face problems at continental and larger scales (e.g., climate change [Schimel et al. 1997] or animal migrations [Bull et al. 2013]), and adequately meeting this challenge requires distributional data produced in a consistent manner, making broad-scale data particularly valuable (Gray and Song 2013; White et al. 2014, Wang et al. 2015). Modeling and analysis methods must, therefore, account for predictor-response variability to avoid spuriously applied relationships (Fink et al. 2010; Ashcroft et al. 2011).

Among various efforts to produce FCC and ISC maps at broad scales, there is little agreement on the effect of local-scale variability in predictor-response relationships. Coulston and others (2012) determined that a single model defined for the entire conterminous United States predicted FCC nearly as well as a locally trained model, but other research suggests an approach sensitive to local variation may be better suited for such broad extents. Homer and others (2007), Kim and others (2014), and Townshend and others (2012) all employed discrete, locally fit models to classify national or global land cover that were later mosaicked to generate a final dataset.

Offering a more seamless approach, Fink and others (2010) developed the spatiotemporal exploratory model (STEM). A STEM is a framework for generating predictions at continental scales, which that study used to model species distributions for the conterminous United States. A STEM aggregates over an ensemble of locally defined, spatially overlapping estimators (i.e., base models). Each pixel in the final prediction map is the average or mode of all overlapping pixels from each estimator, producing a broad-scale map that reflects local-scale variability. Although the main goal of Fink and others (2010) was to develop a model that could effectively predict species distributions with intra-annual spatiotemporal variability, this ensemble approach with estimators limited in spatial extent represents a promising approach for predicting FCC and ISC across broad extents.

The motivation for this study, then, was to develop a broad-scale, fine-resolution land cover model that remains sensitive to local variation. Here, we adapted the STEM algorithm to apply both this novel algorithm and a novel time-series predictor dataset for developing yearly FCC

and ISC maps for California, Oregon, and Washington spanning 1990–2012, a time-series dataset that did not previously exist for such a broad extent and fine resolution. These data will ultimately aid researchers and conservation practitioners in understanding how these influential metrics of land cover patterns affect a wide range of ecological, geochemical, and social processes.

Methods

Study area

We conducted this study using data covering the west coast of the conterminous United States, including all of California (CA), Oregon (OR), and Washington (WA). Although this region does not span an entire continent, the ecological and anthropogenic landscapes together represent the diversity found across a continental extent, making it suitable for testing our methods. Elevations across the study area range from sea level to 4400 m. Both OR and WA are broadly characterized by two vegetation zones, forest and steppe (Franklin and Dyrness 1988), generally corresponding to maritime and continental climatic zones, respectively. Vegetation zones are similarly divided by the southern Cascades and Sierra mountains in California, although more xeric vegetation communities are common in southern California with coastal chaparral in the west (Ashbaugh and Alwin 1994, 110) and desert scrublands throughout much of southeastern California (Miller and Hyslop 1983). Moreover, the entire region exhibits a wide variety of forest types, including wet temperate rainforests along the Pacific coast; highly productive, temperate coniferous forests throughout mid-elevation areas; subalpine coniferous forests; broadleaf forests in frequently disturbed areas and riparian zones; and mixed conifer and broadleaf in southern Oregon and Northern California (Franklin and Dyrness 1988; Ashbaugh and Alwin 1994).

Like the forests throughout the study area, the anthropogenic landscape exhibits broad diversity. Agricultural land dominates much of the low-elevation central valleys throughout the study area, with a wide variety of crop types, including cotton, wheat, vegetables, feed crops, and tropical evergreens (e.g. avocados, citrus trees, olives) (Miller and Hyslop 1983). Across OR and WA, hay, wheat, and potatoes are among the prevailing field crops, while orchard crops make up large proportion of production in WA and grass seed is more common in OR (Ashbaugh 1994, 326–333). Most urban development throughout OR and WA is relegated to major urban centers and highway travel corridors (Ashbough 1994, 382). The same general pattern is found in CA, although development intensity and extent are generally greater across the state. The ecological and population density diversity makes the area ideal for testing this ensemble model, designed to handle globally variable relationships between predictor and response variables.

Overview of methods

Developing our time series of maps first required establishing a properly specified and accurate model of predictor-response relationships. For the STEM algorithm we employed, two parameters constrain a model's ability to estimate those relationships: the number of samples and the size of locally defined estimators. To ensure our model accurately estimated predictor-response relationships, we first tested multiple numbers of samples. With the best sample size, we tested three sizes of locally defined estimators. Using both the best performing sample size and estimator size, we then compared the STEM algorithm to a bagged decision tree, the most comparable model with globally defined estimators. In our last test, we used the best performing STEMs to generate the complete time series and tested the temporal extensibility of the models.

For evaluating the models, we primarily relied on confusion matrices and several derived summary statistics. We additionally developed several maps to inform per-pixel predictions. In summary, our methods involved 5 steps for developing and evaluating our models: 1) testing multiple sample sizes; 2) testing multiple sizes of estimators; 3) comparing our STEMs to conventional models with globally defined estimators; 4) evaluating the time series; and 5) developing per-pixel evaluation maps. Details of these steps are described below.

Training data

Our training data came from the 2001 National Land Cover Database (NLCD) maps. These maps were constructed with predefined discrete modeling regions, generally corresponding to ecoregion boundaries (Homer et al. 2004). The response variable in that study was derived from high-resolution orthophotos, which were classified as either a forest/non-forest or impervious/non-impervious surface and resampled to a 30 m resolution to produce percent cover estimates (Homer et al. 2007). To develop NLCD maps, these percent cover estimates were then related to Landsat composites from imagery ranging from 1999–2002, finally resulting in maps that range from 0–100% cover. Reported accuracy from ten-fold cross-validation for each modeling region across our study area ranged from 79–91% for FCC and 83–93% for ISC (ibid). Using these data for training assumed that NLCD maps represent accurate cover estimates, although we recognized that misclassified pixels introduced error into our training data. While independent assessments have revealed errors in NLCD FCC and ISC data (Greenfield, Nowak, and Walton 2009; Nowak and Greenfield 2010), NLCD maps are considered to be high quality products and are widely used in ecological modeling (e.g., Fink et al. 2010; Pidgeon et al. 2007; Theobald et al. 2009; Wickham, Wade, and Norton 2014; Wilson, Lister, and Riemann 2012).

For accurate classification from remotely sensed spectral imagery, different materials must be separable by their appearance in different portions of the electromagnetic spectrum—that is, their spectral signatures must be distinct. Non-forest and non-impervious pixels occur with a wide range of spectral signatures, so we needed more samples from this class to capture the variability of such occurrences. To ensure training samples adequately represented the full range of percent cover values, we used random samples stratified by decile with an additional stratum reserved for 0-value (i.e., non-forest or non-impervious) pixels. Our intent was to offer each estimator enough information to accurately predict for the various situations that might result in non-forest or non-impervious pixels, while avoiding oversampling 0-value pixels that may be misclassified or for which there could be an error in either the predictor or training data. Still, providing enough 0-value samples for each estimator to adequately handle such variability is a known challenge for broad-scale FCC and ISC predictions (Huang et al. 2001).

Predictor data

Spectral predictor data were all products of LandTrendr, a tool for developing time series images and landscape change maps from the Landsat archive (Kennedy et al. 2012). LandTrendr is unusual among land cover change detection algorithms in that it employs temporal segmentation, which allows for detection of both abrupt changes (i.e., short, fast disturbances such as wildfire) and gradual trends (i.e., long, slow disturbances like insect damage). The spectral trajectory of each pixel is idealized as a series of fitted straight-line segments, which omits noise while retaining essential detail. For every year in the time-series stack, fitted images are created by interpolating each pixel's value to the appropriate time step along these segments. Each pixel is

ultimately fitted to its own trajectory (Kennedy et al. 2015). LandTrendr outputs we used in this study included fitted tasseled-cap brightness, greenness, and wetness; normalized burn ratio (NBR); change from the previous year in brightness, greenness, wetness, and NBR; the number of years since the last detected disturbance; and the mean squared error of each yearly image.

Other predictors we used included elevation and several derivatives from a 30-meter digital elevation model (Gesch et al. 2002). Elevation derivatives included slope represented as percent and cosine-transformed aspect (Beers, Dress, and Wensel 1966). Aspect represented in this way is useful because the maximum value represents the maximum distance in data space from the minimum value, whereas the same is not true for aspect ranging from 0° to 360° .

Model description

The STEM algorithm we used is a discrete ensemble model, composed of a population of independent estimators. As a general class, ensemble models aim to improve predictions by aggregating over numerous estimations using a common base model algorithm. Each estimation “represents a possible state that the real system might be in at some specified time” (Araújo and New 2007). All ensemble methods can be broken down into two separate tasks: establishing a population of estimators, and combining those estimators to determine the final prediction (Hastie et al 2009, 605).

The spatiotemporal boundaries of estimators are delineated by support sets, which define the spatial and temporal domains of each estimator. For most ensemble models, each support set spans the entire spatial extent of the modeling region, offering global spatial support to each

estimator. To avoid long-range learning from spurious relationships at broad extents, however, support sets of a STEM are all limited in spatial and temporal extent, establishing only local support for each estimator. Support sets are geographically stratified, randomly located, and spatially overlapping (see Figure 2.1). Each estimator of the STEM produces predictions for the spatiotemporal extent defined by its support set. Because support sets overlap, the value of each pixel at a fixed geographic location is estimated multiple times by different estimators. A final prediction is determined by calculating either the mean or mode for each spatiotemporally overlapping pixel. This local aggregation reinforces both local scale homogeneity and global scale heterogeneity (Fink et al 2010). In this study, we calculated both the mean and the mode of all overlapping pixels and compared the resultant predictions. The temporal extent of estimators in our STEMs was fixed by using training and predictor data only from 2001 for developing the model.

The STEM framework is intended to be flexible, allowing for any base model algorithm to predict within support sets. We elected to use a decision tree as our base model algorithm for several reasons. Decision trees are a nonparametric machine-learning method designed to efficiently partition a large number of predictors to develop classes with the greatest discrimination between each class, according to a given information criterion (Breiman 1984, 56; Fink et al. 2010). The training data are recursively split, building a tree-like structure. One commonly cited weakness of decision trees is their tendency to overfit training data (Bramer 2002). While an overfit tree favors specific portions of the data space, an ensemble of overfit trees can potentially fill and adequately represent the full data space (Dietterich 2000).

Aggregating over a large population of estimators can have a similar effect to other methods of

generalizing trees (Breiman 2001); thus, an aggregated population of overfit trees is one method to achieve an optimal prediction.

Another known weakness of decision trees is their instability (Breiman 1998; Opitz and Maclin 1999). That is, predictions are highly susceptible to changes in the training data (Breiman 1996b). Breiman (1996a) introduced the concept of bagging (*bootstrap-aggregating*) to stabilize predictions for such sensitive procedures. Like other ensemble methods, bagging combines multiple, slightly varied estimators. Each of these is trained on a bootstrap sample of training data, i.e., sampled with replacement (Johnson 2001). Bagging can potentially improve performance for unstable base models like decision trees, as each estimator receives limited information about the data space. This ultimately produces a highly diverse population of decision trees. Indeed, the strength of a bagged model relies on the diversity of the ensemble (Dietterich 2000), as each member may favor different predictor-response relationships. It is reasonable to assume, then, that bagging would improve predictions at broad scales if relationships between predictors and a response variable vary across the modeling region. Thus, we used bagged estimators to more effectively estimate these relationships.

Evaluating map agreement

The general goal of map evaluation is to compare a sample of predicted values to values believed to represent reality. In our case, we evaluated our predictions against NLCD maps at locations not used to train the models. Following the recommendations of Foody (2002), we employed assessment methods that describe the types, spatial distribution, magnitude, and frequency of error. Riemann and others (2010) and Strahler and others (2006) insist that an adequate

assessment should also include multiple evaluation methods, as each may provide different information. We primarily relied on confusion matrices with map classes defined by decile and summarized them with several different statistics to assess agreement between STEM and NLCD maps. We calculated user's and producer's accuracy for each stratum, and overall accuracy and Cohen's kappa coefficient (1960) to summarize the agreement for the entire test sample set. Kappa compares the observed overall accuracy to the expected accuracy, taking into account the random chance of agreement between a reference map and a prediction map. Although kappa is a common metric used to evaluate land cover classifications, its efficacy has been questioned (Foody 1992; Liu, Frazer, and Kumar 2007; Pontius and Millones 2011; Oloffson et al. 2013). We therefore calculated two alternative metrics of disagreement recommended by Pontius and Millones (2011): quantity disagreement and allocation disagreement. Quantity disagreement can be understood as the difference in the amount of each class in a prediction map as compared to a reference map. Allocation disagreement considers the number of times a map class is misplaced. Both quantity and allocation disagreement range from 0 to 1, where 0 represents perfect agreement.

In the process of assessing our models, we discovered pixel misalignment errors of up to one pixel in the source LandTrendr data. Because these data were processed by Landsat scene, the offset varies by scene (Figure 2.2), preventing correct alignment of the source data before training and predicting. We compensated for this error by adjusting our sampling scheme when generating both training and test samples. Our training sample values were derived from the average of a 3 x 3 pixel kernel centered on each randomly generated sample. For evaluating the models, we used two different sampling schemes using a 3 x 3 kernel also centered on testing

samples: 1) the average of the nine reference and prediction values (hereafter, “average” method), and 2) the pixel of the reference data with the smallest difference from each center pixel test sample (hereafter, “best-match” method). By using the best-match pixel, it is more likely that the pixel-to-pixel comparison necessary for map scale assessments would be performed on the correct pixels. Results from both methods are reported below.

Sample size

The selection of an appropriate sample size for a modeling task can have a prominent effect on prediction accuracy (Foody 2002), as training samples must represent the full range of variability in predictor-response relationships. Too much of the same information, moreover, can produce an overfit model, a common occurrence with decision trees (Bramer 2002). While statistical methods exist for estimating an appropriate sample size, the number required varies by prediction task and modeling method. Therefore, we tested the effect of sample size on multiple models trained on samples of varying sizes from 100,000 (< 0.1% of total pixels) to 1,500,000 (~1.5% of total pixels) total samples. For each sample size, we trained the model on 80% of the total samples and reserved 20% for model evaluation.

Support set size

Optimal support set size is a balance between limiting the complexity of predictor-response relationships while still offering enough information for each estimator to make adequate predictions. Ultimately, the optimal size is a reflection of the scale important to the response variable—that is, the distance beyond which learned relationships from one location are no longer valid. Since that distance is unknown for FCC and ISC in our modeling region, we tested

three support set sizes for both FCC and ISC: 200 x 100 km (small), 400 x 300 km (medium), and 800 x 600 km (large) (see Figure 2.1).

Comparison of a STEM to a conventional bagged decision tree model

Since reasonable expectations of accuracy vary by modeling task and research objectives (Foody 2008), the value of any new modeling method is best compared to a next-best method (Pontius and Millones 2011). Using the sample size and support set size that produced the best performing model we compared STEM predictions to predictions made with a bagged decision tree (BDT), the most comparable bagged ensemble model with global support for all estimators. A BDT is essentially a STEM but with all support sets covering the entire modeling region.

Temporal extensibility

As described above, the purpose of this study was not only to demonstrate the utility of a STEM for creating FCC and ISC maps, but also to leverage a novel yearly time series dataset to produce temporally dense FCC and ISC time series data as well. In our last test, we used the best performing FCC and ISC models developed with 2001 NLCD training data and predicted both variables for all years from 1990 to 2012. To evaluate the temporal extensibility of the model, we also compared 2011 STEM predictions to 2011 NLCD maps, the only other year of NLCD data with both FCC and ISC maps.

Spatial assessment of error, uncertainty, and variable importance

In addition to using measures of map agreement, we evaluated model error, uncertainty, and explored prediction results with three per-pixel assessment methods. Foody (2002) and Riemann

and others (2010) note the importance of demonstrating the spatial variability of model accuracy, as a data user may only be interested in a particular area of a prediction map. Such specific use is even more likely for broad-scale, fine-resolution datasets.

First, we introduce a novel method to evaluate model performance across the modeling region and to illustrate the spatial variation in predictive performance. Our evaluation method is a modification of a common assessment strategy for bagged models, the out-of-bag (OOB) score. For a typical bagged model, the OOB score is calculated by assessing the error of each estimator. Since each estimator is trained on a random subset of the training data, a prediction is made with the training samples not used to train that estimator, and the classification accuracy, or OOB accuracy rate, is calculated. The OOB score for a typical bagged model is the average of the OOB accuracy rates of all estimators (Breiman 2001), ranging from 0 to 1, where an OOB score of 1 indicates a perfect classifier (Breiman 1996b). For our STEMs, we still determined OOB accuracy per estimator, but each estimator was limited in spatial extent by its support set, so the average OOB accuracy varied across the modeling region. Leveraging this variability, we averaged all overlapping accuracy values to determine the OOB score for each pixel, resulting in an OOB score map that illustrates how accuracy varies across the modeling region (Figure 2.3). We summarized the overall OOB score for the whole STEM by averaging all pixel values.

Uncertainty is distinguished from error in that it is a measure of the confidence in a prediction rather than simply the difference between known and predicted values. For decades, researchers have called for more explicit assessments of uncertainty and spatial assessments of uncertainty, in particular (MacEachren 1992; Foody and Atkinson 2002). Spatially explicit information on

uncertainty can provide crucial information on proper use of a map. Since bagged models rely on a population of independent predictions, they provide a straightforward platform for calculating uncertainty for each pixel (Sanyal et al. 2010). In the same way that we aggregated predictions per pixel, we calculated the standard deviation of all spatially overlapping pixels. These standard deviation maps provide critical information for interpreting prediction results.

Another advantage of decision-tree-based ensemble approaches is that, in the process of constructing such models, the relative importance of each predictor is easy to compute. Relative importance is determined per estimator by calculating the increase in information gain (i.e., the increase in the model's discriminatory ability) after including that predictor in the model. For each predictor variable, importance scores are averaged for all estimators to determine importance for the entire ensemble. Although there are some limitations to calculating variable importance with this method (Strobl et al. 2007), importance remains an informative byproduct of decision-tree-based models. Moreover, a typical ensemble model can only assess variable importance for the entire extent of the modeling region. With locally fit estimators, however, we estimated variable importance per pixel in much the same way that we calculated per-pixel OOB scores and standard deviation. Here, we determined the most important predictor variable for each estimator and aggregated by mode to approximate the most important variable per-pixel.

Results

Overall model performance

In general, FCC and ISC STEMs performed well, with overall accuracies of 89.6% and 98.0%, respectively. Kappa scores also indicated general agreement between our models and NLCD

maps, with FCC scoring 0.82 and ISC scoring 0.71. Per-stratum accuracy metrics for both FCC and ISC followed a similar pattern throughout all model iterations. For the best performing FCC model (Table 1), user's and producer's accuracies were highest for the 0–10% value stratum, with a user's accuracy of 97.6 % and a producer's accuracy of 98.2%. User's accuracy was lowest for the 11–20% stratum at 46.8%, and producer's accuracy was lowest for the 21–30% stratum at 63.3%. As with FCC, the best performing ISC model (Table 2) yielded the highest user's and producer's accuracies in the 0–10% stratum at 99.6% and 99.1%, respectively. The lowest user's and producer's accuracies occurred in the 11–20% stratum at 49.9% and 35.9%, respectively. Overall, the models struggled more to predict mid-range values than either extremely low or extremely high values.

Predictions aggregated by the mode of all overlapping pixels consistently agreed with the reference map more than the predictions determined by the mean. The overall accuracy of FCC mode predictions for all models was approximately 7% higher than FCC mean predictions, and the kappa was consistently greater by 0.06–0.07. The accuracy of ISC mode predictions for all models exceeded mean predictions by 5–11% for overall accuracy and 0.07–0.14 for kappa. The decrease in performance was largely attributable to commission errors in the 0–10% stratum. Mean predictions consistently predicted high impervious values in areas where no impervious pixels exist in the NLCD reference map. As a result, producer's accuracies for mean predictions for this stratum were as much as 11% lower than the mode predictions. Because the mode predictions yielded consistently better results, all model results reported here are from those maps and we used only the mode predictions to create the time series.

The best-match sampling method yielded considerably better results than the average method, as demonstrated by the narrow, concentrated plots of two-dimensional histograms (Figure 2.4). This is to be expected, considering that the average method inevitably drew from pixels in both the reference and the prediction maps that did not represent the same locations on the ground. We ultimately adopted the best-match method for evaluating model parameters and for other tests. All results discussed below are, therefore, drawn from the best-match method.

Sample size

Most evaluation metrics improved nearly monotonically with increasing sample sizes for both FCC and ISC (Table 3). In general, sample size had a greater influence on ISC models than FCC, as overall accuracy, kappa, and OOB scores for FCC varied less than for ISC. All accuracy metrics show that FCC and ISC models trained with oversampled 0-value pixels performed better than models trained with an even number of samples in all strata. The best-performing sample size for both models was 1,500,000 (the largest tested) with ten times the number of 0-value pixels as all other strata.

Support set size

For both FCC and ISC, model performance decreased as support set size increased, although this was clearer in the FCC models than ISC. For FCC models, this effect was demonstrated by the accuracy, kappa, and average OOB score, but for ISC it only affected the average OOB score and kappa. We attribute this lack of response in ISC accuracy to the uneven distribution of map values. For all support sizes, ISC user's and producer's accuracies in the 0–10% stratum were relatively constant, with differences of $< 0.2\%$ between support set sizes. This map stratum also

covers over 95% of the modeling region, so it has an overwhelming influence on overall accuracy. For nearly all other map strata, however, the smaller support size outperformed the larger with an increase in producer's accuracies between 0.2–10% and an increase in user's accuracies between 0.3–2.6% (see Appendix A). Since kappa accounts for the random chance of agreement, it gives greater consideration to the accuracy of these rare classes than overall accuracy does. Thus, the uneven distribution of performance in different strata had a clear effect on kappa, but it was not reflected in the overall accuracy. We, therefore, used the smallest support set size for comparing to BDTs and generating the time series.

Comparison of a STEM to a BDT

As anticipated, both FCC and ISC STEMs with locally defined estimators outperformed BDTs with globally defined estimators across all evaluation metrics (Figure 2.5). For the best performing FCC STEM, we achieved an overall accuracy of 89.6% with a kappa coefficient of 0.82, a quantity disagreement of 0.01, an allocation disagreement of 0.09, and an average OOB score of 0.80. In comparison, the FCC BDT produced an overall accuracy of 84.9% with a kappa coefficient of 0.74, a quantity disagreement of 0.02, an allocation disagreement of 0.13, and an OOB score of 0.73. For the best performing ISC STEM, we calculated an overall accuracy of 98.0% with a kappa coefficient of 0.71 and an average OOB score of 0.85. The ISC BDT showed an overall accuracy of 97.2% with a kappa coefficient of 0.65 and an average OOB score of 0.79. Both models achieved the same quantity and allocation disagreement scores of 0.00 and 0.02, respectively.

Geographically, the FCC BDT tended to over-predict percent cover in wet forests (the Coast Mountains in northwestern CA and southwestern OR) and under-predict in dry forests (e.g., Sierra Nevada Mountains in CA). The ISC BDT model generally over-predicted around some urban centers like Seattle, WA and Los Angeles, CA and under-predicted in agricultural areas. For STEMs the mean squared error predictor helped improve predictions only in these areas, but since the BDT needed to learn predictor-response relationships for the whole modeling region, the information offered by the mean squared error was overshadowed by more prominent relationships. Difference images between STEMs and BDTs illustrate these spatial patterns of BDT performance (Figure 2.6).

Temporal extensibility

After determining the best sample size, support set size, and aggregation method, we used the best performing model to create maps for the entire time series from 1990-2012. Evaluating our 2011 ISC and FCC maps with 2011 NLCD data, our ISC prediction performed nearly as well as our 2001 prediction with overall accuracy of 97.8% (0.3% lower than 2001) with a kappa score of 0.69 (0.02 lower than 2001). The 2011 STEM FCC prediction, however, exhibited markedly less agreement with 2011 NLCD reference data. That prediction, when compared to NLCD, yielded an overall accuracy of 76.2% (13% lower than 2001) and a kappa coefficient 0.61 (0.21 lower than 2001). The potential reasons for the difference in 2011 prediction accuracy between FCC and ISC are discussed below.

Spatial assessment of error, uncertainty, and variable importance

While map-scale assessments provided information on overall model performance, prediction accuracy and uncertainty inevitably varied across the modeling region. We assessed this variability by calculating per-pixel OOB scores and standard deviations (Figure 2.7). For both FCC and ISC, the highest OOB scores coincided with areas consisting of mostly 0-value pixels, with some support sets achieving an OOB accuracy rate of 1. Where the majority of pixels had a value of 0, the predictor-response patterns within estimators were far less complex. Although these 0-value pixels likely represent a wide range of actual materials, most of the training samples all corresponded to a single percent cover value so the spectral differences were inconsequential. In comparison, FCC and ISC predictor-response patterns in forested and developed areas were much more complex, potentially with different materials represented by the same percent cover value. The OOB maps show that estimators in these areas struggled more to classify FCC and ISC correctly (Figures 2.7c and 2.7d). The lowest OOB scores for FCC were west of the Cascade Range in OR and WA, with a prominent depression in southwestern OR and northwestern CA. Major depressions in ISC OOB scores were centered on Seattle, WA and Sacramento, CA with a moderate depression around San Diego, CA.

Standard deviation maps (Figures 2.7a and 2.7b) complimented the OOB maps, providing crucial information on ensemble prediction structure. Standard deviation tended to be highest for pixels representing materials that are spectrally similar to either forests or impervious surfaces. For FCC, standard deviation ranged from 0-47% cover with a mean of 4.76%. The highest standard deviations were most common in agricultural areas, which bear strong spectral resemblance to trees. High standard deviations were also found in forest subject to logging and other high

magnitude, high frequency disturbances. FCC standard deviation was only moderate, however, in densely forested areas less subject to disturbances, such as areas in and around Olympic National Park in WA and higher elevation forests in WA, OR, and CA. The lowest FCC standard deviations consistently occurred in non-forested areas such as the deserts east of the Cascade Range and in southern CA. Standard deviation for ISC also ranged from 0-47% cover but with a mean of 3.44%. Standard deviation values were highest in areas dominated by agriculture, such as the Willamette Valley in OR, much of central WA, and the Central Valley, Salinas River valley, and Coachella Valley in CA. Standard deviation was also high in bare ground desert areas, particularly near population centers. Standard deviation was lowest in regions far from population centers and areas dominated by forest, a dark land cover that is easily contrasted with bright impervious surfaces.

Perhaps the most useful spatially explicit method for understanding model dynamics that we employed was calculating variable importance maps (Figure 2.8). In general, importance maps demonstrated that predictor-response relationships differ regionally and that STEMs were sensitive to this local-scale variability. Importance maps additionally revealed several unexpected relationships between predictor and response variables, which we discuss in greater detail below.

Discussion

Our ultimate goal in this study was to develop dense time series FCC and ISC datasets at a broad scale. Doing so required assessing the appropriate training sample size by testing different numbers of samples and addressing the wide variability in predictor-response relationships found

at broad scales. The STEM method addresses the latter problem with an ensemble of locally defined estimators, limiting the predictor-response variability per estimator but representing the entire data space (and geographic space) across the whole ensemble. With only local predictor-response relationships influencing the final prediction for each pixel, the STEMs were able to represent local scale homogeneity, while remaining sensitive to global scale heterogeneity. Also recognizing that a novel method is only worth adopting if it improves upon existing methods, we compared our predictions to that of a BDT. Similarly, the time series dataset we developed will only be useful to others if it depicts an equally accurate representation of reality as other comparable products. To test this, we compared our 2011 STEM maps to 2011 NLCD maps. Lastly, we developed additional maps to inform per-pixel predictions and to qualitatively describe results.

Sample size

Selection of an appropriate sample size is a common challenge for any predictive modeling task. After testing increasing sample sizes, we found consistent gains in model performance with each increase and no signs of diminishing accuracy with larger sample sizes. These findings are contrary to Coulston and others (2012), which concluded that a relatively small sample size of 1,000 was adequate for predicting FCC for the conterminous United States, and more samples yielded negligible gains. The methods of that study, however, differ from ours in several important ways. Most importantly, the training data were photo-interpreted samples from high-resolution imagery. The NLCD data we used to train our models, in contrast, were modeled data that used either a forest or urban mask to assign a value of 0 to any pixels determined to be non-forest or non-impervious. The masks effectively created a highly diverse class of 0-value pixels

representing a wide range of actual materials, some of which were not easily separated spectrally. Indeed, the intent of the original forest and urban masks was to remove commission errors from the NLCD predictions at locations that should have had a percent cover of 0. While we observed increased overall performance with larger sample sizes, we saw the largest gains in the lowest map stratum of values ranging from 0–10%. Furthermore, oversampling 0-value pixels proved to be the most effective sampling scheme, presumably because the total training set more adequately represented the variability in predictor-response relationships. This underscores the point that an appropriate sample size is one that sufficiently represents the entire data space. The exact number of samples necessary to achieve that sufficiency will inevitably vary with different training data, predictor data, and modeling tasks.

Another important difference contributing to disparities between our results on sample size and that of Coulston and others (2012) is the model algorithm used. That study employed a random forest model, a model similar to a BDT with globally defined decision trees, but it is recognized as a marked improvement over the BDT algorithm (Breiman 2001). Each estimator of a random forest is defined globally, drawing support from the entire modeling region. With this global support, each estimator shares more of the same training samples, requiring a relatively small total number of samples. With the limited spatial extent of each estimator for a STEM, however, fewer samples are shared across each estimator. Since each one requires an adequate representation of the data space, a STEM demands a larger total number of samples; the smaller the support sets are, the larger the total number of samples must be. This is one potential weakness of the STEM algorithm, as a large training set may not be available depending on the source of the training data.

Size of support sets

The size of support sets of a STEM controls the potential variability of predictor-response relationships within an estimator. The degree to which support set size affects a STEM is largely dependent on the predictor-response variability across the entire modeling region and the scale of those relationships. The scale important to a phenomenon is considered small if those relationships change over short distances. Expecting a wide range of predictor-response variability, we anticipated that a model more sensitive to small-scale change would be most appropriate for predicting FCC and ISC in our study, as CA, OR, and WA collectively represent considerable ecological diversity.

As expected, we generally found decreasing model performance for both FCC and ISC with increasing support set size. Although this effect was subtle, it suggests that small-scale predictor-response relationships exercise primary influence. FCC models exhibited this trend more clearly, with a monotonic decrease in accuracy across all evaluation metrics. As support set size grew, so did the complexity of predictor-response patterns contained within each support set. The corresponding estimators struggled to tease out differences in spectral signatures of different materials, ultimately leading to misclassification.

While set size had a similar effect on ISC models, the results were more nuanced. The average OOB score and kappa coefficient decreased with increasing set size, but the overall accuracy remained constant for all sizes. This varied response of accuracy metrics was the result of uneven accuracies across different map strata and the differently sized areas that those strata represent.

As described above, medium and large set models struggled with higher ISC values, showing frequent commission errors. This is largely due to the spectral similarity of impervious surfaces to some non-impervious surfaces such as bare ground. Urban areas with high concentrations of impervious surfaces (i.e., high ISC values) are highly localized across the study area. With small support sets, estimators that at least partially covered these dense concentrations of impervious surfaces needed to also predict for only a small number of low-value ISC pixels, limiting confusion of non-impervious surfaces with impervious surfaces. Larger support sets, however, forced each estimator to learn predictor-response relationships with materials that were not easily separated by spectral characteristics. Put more simply, smaller support sets resulted in less complex predictor-response patterns per estimator, so each individual ensemble member produced more accurate predictions. We only tested three support set sizes, so the optimal support set size of a STEM remains a topic for further study.

It is important to note that decreasing support set size may not always produce better-specified models. The optimal size likely depends on both the modeling task and the spatial distribution of training data. The size of support sets is a rough approximation of a scale important to the phenomenon; greater model accuracy with smaller support sets indicates that small-scale predictor-response relationships predominate, but for phenomena with lower spatial frequency smaller support sets may not offer each estimator enough information to adequately make predictions. The same may be true if training data are unevenly distributed across the modeling region. Fink and others (2010) used a STEM to model intra-annual bird distributions with spatially biased data. That study concluded that larger support sets increased model performance, as estimators in areas containing sparse training samples may not have received enough training

samples to make accurate predictions. Independent exploration of the appropriate support set size of a STEM is, therefore, necessary for any new modeling task.

STEM vs. BDT

The BDT models we developed are essentially global-support STEMs, as all estimators cover the entire extent of the modeling region. Thus, the patterns of differences in accuracy between the BDTs and the small set-sized models are even more evident than with medium and large set-sized STEMs. For instance, user's and producer's accuracy for the 0–10% strata of ISC BDT models were only 0.3% and 0.2% lower, respectively, than the small support set STEM. For all other strata, the average difference in user's and producer's accuracy was > 7%. The same pattern exists in per-stratum accuracy for FCC models. Here again, the support from broad extents presents each estimator with predictor-response patterns that are too complex to adequately learn. The decrease in all other accuracy metrics between small-set STEMs and BDTs further illustrates the poor performance of models with broad-extent estimators. Interestingly, these results comparing STEMs to BDTs contradict the findings of Coulston and others (2012), a pilot study that developed the methodology for creating 2011 NLCD FCC data. That study concluded that a single model for the entire conterminous United States only slightly underperformed locally defined models. The differing results may be a product of either sampling intensity (discussed above) or model algorithm (discussed below).

Temporal extensibility

The ultimate goal of this study was to develop yearly maps of FCC and ISC, two yearly time series datasets that did not previously exist at a broad scale. We assessed the temporal

extensibility of our models by comparing our predictions to 2011 NLCD FCC and ISC maps, the only comparable FCC and ISC maps that overlap spatially and temporally with our modeling region. Our 2011 ISC map showed only a slight decrease in agreement with 2011 NLCD as compared to 2001 maps, but our FCC data showed a considerable drop in agreement. Here again, the difference is attributable to methodological disparities between our study and that of Coulston and others (2012). The random forest model and sampling scheme used to create the 2011 NLCD FCC maps departed so significantly from methods used to create the 2001 NLCD FCC data that the authors concluded they could not adequately compare their results with that of the 2001 effort (*ibid*). Our 2011 predictions are similar in character to 2001 NLCD maps, since we relied on 2001 NLCD data to train our models. Additionally, we found that predictions aggregated by mode produced maps that agreed most with 2001 NLCD maps, so we also aggregated all time series predictions by mode. The 2011 NLCD FCC map was aggregated by mean, which dramatically reduces the number of low- and high-value pixels. To make a more appropriate comparison with 2011 FCC NLCD maps, we also calculated the mean prediction for the 2011 STEM FCC model. The differences in maps aggregated by mean and mode are evident in a comparison of cumulative distribution functions between NLCD, STEM mode, and STEM mean predictions (Figure 2.9). All three maps agree most in the middle of the distribution, but the NLCD and STEM mean contain far fewer pixels with $FCC < 20\%$ and $> 90\%$ than the STEM mode. Overall, the STEM mean shows greater agreement with NLCD than the STEM mode, which is to be expected considering the greater similarity in methods. Whether or not the mean or the mode better represents actual FCC—or ISC, for that matter—can only be determined with an independent test sample.

Spatial assessment of error, uncertainty, and variable importance

All predictive models carry inherent error and uncertainty, as models are only representations of real systems. This can be particularly true of maps created from reference data that are also modeled rather than directly measured on-site, a common practice in the development of geospatial data. Although all data collection methods introduce some error, error in modeled data can be markedly larger. Errors also are usually not evenly distributed, particularly for spatially explicit broad-scale models. Spatially quantifying both error and uncertainty are, therefore, integral to adequate map assessment (Foody 2002, Riemann et al. 2010) to ensure map products are used and interpreted appropriately.

In our study, both training and predictor data were sourced from modeled predictions, and both potentially introduced error to the modeling process. Nowak and Greenfield (2010) assessed 2001 NLCD FCC and ISC using independently derived estimates from manual photo interpretation. That study found up to a 28% difference between predicted FCC and photo-interpreted estimates and up to 6% difference in ISC estimates. 2001 NLCD FCC and ISC products also include estimates of error obtained from ten-fold cross-validation with average errors per modeling region between 9–21% for FCC and 7–17% for ISC (Homer et al. 2007). Additionally, most of the predictor variables in our study were products of LandTrendr, a tool for modeling yearly time series images and extracting information about landscape change. Smoothing the time series is essential for reducing noise, but it can be difficult to distinguish between noise and true signal. The error introduced in the smoothing process can be represented as a map with the mean squared error of each pixel. We used this map explicitly as a predictor to help estimators distinguish between predictor patterns from materials that may have otherwise

looked similar in the data space. This error map also indicates potential areas of uncertainty in the final predictions. Both the 2001 NLCD errors per modeling region and the LandTrendr mean squared error are shown in Figure 2.10.

Another potential source of error from LandTrendr data in our FCC and ISC predictions is the pixel misalignment discussed above. We adjusted our sampling schemes for generating training and testing data to account for this inherent error in the predictor data. The differences in our results between the average and best-match sampling scheme indicate that the best-match method sufficiently handled the misalignment. This may have led to overly optimistic assessments, however, as the best-match sampling scheme may not have always produced a match with what would have been the correct pixel if the data were properly aligned.

To spatially assess the error in our maps, we implemented a modification of the OOB score, leveraging the limited spatial extent of estimators in a STEM. The result of this was an OOB map with per-pixel approximations of accuracy (Figures 2.7c and 2.7d). While OOB scores corresponded positively with complexity in predictor-response relationships, artifacts of the training data and the process used to develop them may have also contributed to the patterns in these maps. Nowak and Greenfield (2010) found the highest difference between 2001 NLCD FCC maps and photo-interpreted points in southwestern OR and northwestern CA, the area with the lowest FCC OOB scores for our models. NLCD cross-validation error for this region was not reported in its metadata so it is unclear if that error could have contributed. For ISC, the lowest OOB scores were centered on Seattle, WA and Sacramento, CA with a moderate depression around San Diego, CA. In this case, it is conceivable that these depressions are an artifact of how

the training data were processed. 2001 NLCD data relied on imagery spanning from 1999–2002 (Homer et al. 2007), so depending on cloud cover or image availability, different pixels may have represented different years. Our time series shows some increasing development during this time period, but it is likely not enough to explain the magnitude and highly localized nature of error in these areas. Either way, these more dramatic low OOB scores are inconsistent across population centers, more suggestive of locally driven causes than systematic prediction errors.

To represent the uncertainty of each per pixel FCC and ISC estimate, we calculated the standard deviation of all overlapping pixels from different support sets (Figures 2.7a and 2.7b). Much like the OOB maps, standard deviation was highest in areas with more complex predictor-response relationships. Standard deviation was also high for pixels representing materials that were spectrally similar to either forests or impervious surfaces. For FCC, this was mostly in agricultural areas where vegetated fields bear strong spectral resemblance to forests. Other areas of high standard deviation included frequently disturbed forests, such as those managed for timber production. Here too, multi-year composite imagery used to develop NLCD maps may have confused the signal that indicated what should have been low or high FCC in these areas. For ISC, materials that were not easily separated spectrally included most bare ground desert areas, but high standard deviations were more common for bare ground pixels near large population centers. The estimators in these areas likely received training samples representing both impervious surfaces and bare ground desert, two materials that are difficult to separate from their spectral signals. This also included agricultural areas because many agricultural plots may go fallow in a given year or are only partially covered by crops, exposing large areas of bare ground that appear bright like impervious surfaces. Bare ground desert pixels outside of

developed areas generally had low standard deviations, since the estimators in those areas would not have received the confusing signal of high ISC in the training data. Pixels representing some large bodies of water also had high standard deviations. Not only do these standard deviation maps provide useful information for future data users, they also inform prediction results and provide transparency for interpreting STEM dynamics.

Variable importance maps (Figure 2.8) also revealed interesting and unexpected relationships between predictor and response variables. In western OR and WA, for instance, it was wetness—not greenness—that was the most important predictor of FCC. While forests are overwhelmingly green, so are many agricultural areas. Often agricultural fields that require irrigation are still dry, though, so separating green forest cover from green agricultural cover relied more on wetness. This same relationship is found in the agricultural area south of the Salton Sea near the CA-Mexico border. Only in the desert of southeastern CA, where there is relatively little agriculture and other non-forest vegetation, was greenness most important. East of the cascades and in the Central Valley of CA, brightness was the most salient FCC predictor. Here again, agricultural areas could not be easily differentiated from forests by greenness alone, but rather the darkness of forest contrasted with the brightness of bare ground and distinguished these materials best. NBR was otherwise the most important predictor.

ISC importance maps revealed similarly interesting relationships. For much of the northern half of the study area, greenness was the most salient predictor, presumably because most of the non-impervious areas surrounding development were green and impervious surfaces were not. Much like greenness could not distinguish forest from non-forest, impervious surfaces share brightness

characteristics with other non-impervious materials. In drier areas and a few other local regions, elevation was most important. The reason for this is likely that development is most commonly in low-lying areas. In some mountainous regions, estimators relied on slope for a similar reason. In desert regions of southeastern CA and southeastern OR, mean squared error provided the greatest information gain. In these areas, estimators used the error in LandTrendr data to separate impervious surfaces from similar surrounding desert bare ground. The most intriguing relationship in the ISC importance map is also in southern CA near the Salton Sea. Here is the only place where the change in wetness from the previous year was the most salient predictor. Around the edges of the Salton Sea and in the surrounding agricultural fields, there was marked change in wetness between 2000 and 2001. Without considering wetness from the previous year, pixels that were dry in 2001 would have otherwise resembled adjacent impervious surfaces. Impervious surfaces, however, would not have been wet in the previous year, so estimators could distinguish these materials with the additional landscape change information. Change in wetness presumably influenced predictions in areas with similar landscape dynamics, but it was not the most important anywhere else. Interestingly, wetness was the third most salient predictor for the entire model, but it was never most important for any single pixel. This further illustrates the exploratory value of an ensemble of locally defined estimators for understanding these complex spatial and temporal landscape dynamics.

Conclusion

With increasing awareness of large-scale ecological and social problems, the demand for data that are produced at both broad scales and fine resolutions has only increased. Making predictions at broad scales, however, presents unique modeling challenges, as we have

demonstrated here. A model must be able to address the full range of complexity in predictor-response relationships, and this complexity increases with the extent of a modeling region. If this can be executed successfully, continental or global scale models produced in a consistent manner but that remain sensitive to local patterns can be useful for a multitude of other applications including other broad scale analyses or comparing phenomena in geographically disparate areas.

We have shown that the STEM approach is not only appropriate for developing broad scale land cover maps, but also that it outperforms a typical bagged ensemble that relies on globally defined estimators. Overall, we found that the more localized estimators were, the better the entire ensemble performed. Furthermore, our results using these models in conjunction with LandTrendr data to generate time series are encouraging. Together, this novel algorithm and predictor dataset enable accurate representation of both spatial and temporal patterns.

While we introduced two methods that leverage the local expression that defines a STEM, OOB maps and variable importance maps, there are always potential improvements for any novel modeling approach. We implemented a STEM bearing strong resemblance to a BDT, but other decision-tree-based ensemble models have proven more effective than BDTs. Random forests, for instance, are designed to avoid overfitting, which is still a concern with BDT-like STEMs such as ours. Random forests achieve this by randomly subsetting predictor variables when determining splits of a decision tree (Breiman 2001), an approach that may yield similar benefits as bootstrapping for a STEM. Since the STEM framework allows for any base model algorithm, other base models may be more appropriate for developing land cover maps or for other applications. Support vector machine algorithms, for example, have been used to classify land

cover, with some studies yielding better results than decision trees (Huang, Davis, and Townshend 2002; Shao and Lunetta 2012). These models, however, are known for their stability (Townshend et al. 2012), and bagging is known to actually decrease the accuracy of stable models (Breiman 1996b). Much like the base model algorithm, though, the per-ensemble sampling scheme of a STEM is not necessarily prescribed, meaning bagging is not a prerequisite for the STEM algorithm.

Finally, it is doubtful that ecosystem gradients would ever align with geometrically abstracted boundaries like the rectangles we used to define support sets in our STEMs. Ecoregions are commonly used for the delineation of modeling regions to address this problem (Homer et al. 2007, Coulston et al. 2012). Ecoregions might still not represent ecological reality, particularly because scales and gradients important for various ecological processes are likely different. One potential approach might be to iteratively randomize the delineation of ecoregions to define the limits of ensemble members. This would effectively make the boundaries of the support sets fuzzy, as with a typical STEM, and simultaneously root their delineation and placement in ecological reality. Existing algorithms for automating ecoregion boundaries could easily be modified for this purpose (Hargrove and Hoffman 2004). Fink and others (2014) improved upon the STEM approach by delineating support set boundaries with quadrees. This introduced differently sized support sets enabling a STEM that reflects processes at multiple scales. Either ecoregional or quadtree-based support sets may yield better results for predicting FCC and ISC.

The ultimate goal of any of these modifications to the STEM algorithm would be to tailor the modeling process to more adequately reflect ecological relationships and their spatial variability

in particular. As the need for data produced consistently across broad extents and at fine resolutions increases, these methods may prove useful for creating such data while accurately reflecting ecological reality. The STEM approach explored here already represents a significant step in that direction.

References

- Araújo, Miguel B., and Mark New. 2007. "Ensemble Forecasting of Species Distributions." *Trends in Ecology & Evolution* 22 (1): 42–47. doi:10.1016/j.tree.2006.09.010.
- Arnold, Chester L., and C. James Gibbons. 1996. "Impervious Surface Coverage: The Emergence of a Key Environmental Indicator." *Journal of the American Planning Association* 62 (2): 243–58. doi:10.1080/01944369608975688.
- Ashbaugh, James G, and John A Alwin. 1994. *The Pacific Northwest: Geographical Perspectives*. Dubuque, Ia.: Kendall/Hunt Pub. Co.
- Ashcroft, Michael B., Kristine O. French, and Laurie A. Chisholm. 2011. "An Evaluation of Environmental Factors Affecting Species Distributions." *Ecological Modelling* 222 (3): 524–31. doi:10.1016/j.ecolmodel.2010.10.003.
- Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. "Aspect Transformation in Site Productivity Research." *Journal of Forestry* 64: 691–92.
- Bramer, Max. 2002. "Using J-Pruning to Reduce Overfitting of Classification Rules in Noisy Domains." In *International Conference on Database and Expert Systems Applications*, 433–442. Springer. http://link.springer.com/chapter/10.1007/3-540-46146-9_43.
- Breiman, Leo, ed. 1984. *Classification and Regression Trees*. The Wadsworth Statistics/Probability Series. Belmont, Calif: Wadsworth International Group.
- . 1996a. "Bagging Predictors." *Machine Learning* 24 (2): 123–140.
- . 1996b. "Heuristics of Instability and Stabilization in Model Selection." *The Annals of Statistics* 24 (6): 2350–83.
- . 1998. "Arcing Classifiers." *The Annals of Statistics* 26 (3): 801–24.
- Budreski, Katherine, Michael Winchell, Lauren Padilla, JiSu Bang, and Richard A Brain. 2016. "A Probabilistic Approach for Estimating the Spatial Extent of Pesticide Agricultural Use Sites and Potential Co-Occurrence with Listed Species for Use in Ecological Risk

- Assessments: Probabilistic Crop Footprint Approach for Use in ESAs.” *Integrated Environmental Assessment and Management* 12 (2): 315–27. doi:10.1002/ieam.1677.
- Bull, Joseph W, Kenwyn B Suttle, Navinder J Singh, and Ej Milner-Gulland. 2013. “Conservation When Nothing Stands Still: Moving Targets and Biodiversity Offsets.” *Frontiers in Ecology and the Environment* 11 (4): 203–10. doi:10.1890/120020.
- Campana, Néstor A, and Carlos E. M Tucci. 2001. “Predicting Floods from Urban Development Scenarios: Case Study of the Dilúvio Basin, Porto Alegre, Brazil.” *Urban Water* 3 (1–2): 113–24. doi:10.1016/S1462-0758(01)00004-8.
- Carlson, Toby N, and S Traci Arthur. 2000. “The Impact of Land Use — Land Cover Changes due to Urbanization on Surface Microclimate and Hydrology: A Satellite Perspective.” *Global and Planetary Change* 25 (1–2): 49–65. doi:10.1016/S0921-8181(00)00021-7.
- Cohen, Jacob. 1960. “A Coefficient of Agreement for Nominal Scales.” *Educational and Psychological Measurement* 20 (1): 37–46. doi:10.1177/001316446002000104.
- Cohen, Warren B., and Samuel N. Goward. 2004. “Landsat’s Role in Ecological Applications of Remote Sensing.” *BioScience* 54 (6): 535–45. doi:10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2.
- Coulston, John W., Gretchen G. Moisen, Barry T. Wilson, Mark V. Finco, Warren B. Cohen, and C. Kenneth Brewer. 2012. “Modeling Percent Tree Canopy Cover: A Pilot Study.” *Photogrammetric Engineering & Remote Sensing* 78 (7): 715–727.
- Defries, R. S., M. C. Hansen, J. R. G. Townshend, A. C. Janetos, and T. R. Loveland. 2000. “A New Global 1-Km Dataset of Percentage Tree Cover Derived from Remote Sensing.” *Global Change Biology* 6 (2): 247–54. doi:10.1046/j.1365-2486.2000.00296.x.
- Dietterich, Thomas G. 2000. “An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization.” *Machine Learning* 40 (2): 139–157.
- Fink, Daniel, Theodoros Damoulas, Nicholas E. Bruns, Frank A. La Sorte, Wesley M. Hochachka, Carla P. Gomes, and Steve Kelling. 2014. “Crowdsourcing Meets Ecology: Hemisphere-Wide Spatiotemporal Species Distribution Models.” *AI Magazine* 35 (2): 19–30.
- Fink, Daniel, Wesley M. Hochachka, Benjamin Zuckerberg, David W. Winkler, Ben Shaby, M. Arthur Munson, Giles Hooker, Mirek Riedewald, Daniel Sheldon, and Steve Kelling. 2010. “Spatiotemporal Exploratory Models for Broad-Scale Survey Data.” *Ecological Applications* 20 (8): 2131–2147.

- Foody, Giles M. 1992. "On the Compensation for Chance Agreement in Image Classification Accuracy Assessment." *Photogrammetric Engineering And Remote Sensing* 58 (10): 1459–1460.
- . 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80 (1): 185–201.
- . 2008. "Harshness in Image Classification Accuracy Assessment." *International Journal of Remote Sensing* 29 (11): 3137–58. doi:10.1080/01431160701442120.
- Foody, Giles M, and Peter M Atkinson. 2002. *Uncertainty in Remote Sensing and GIS*. Hoboken, NJ: J. Wiley. <http://public.eblib.com/choice/publicfullrecord.aspx?p=154955>.
- Franklin, Jerry F., and C. T. Dyrness. 1988. *Natural Vegetation of Oregon and Washington*. Portland, Or., Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Dept. of Agriculture. <http://ir.library.oregonstate.edu/xmlui/handle/1957/34706>.
- Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. "The National Elevation Dataset." *Photogrammetric Engineering and Remote Sensing* 68 (1): 5–+.
- Goward, Samuel N., Jeffrey G. Masek, Warren Cohen, Gretchen Moisen, G. James Collatz, Sean Healey, R. A. Houghton, et al. 2008. "Forest Disturbance and North American Carbon Flux." *Eos, Transactions American Geophysical Union* 89 (11): 105–6. doi:10.1029/2008EO110001.
- Gray, Josh, and Conghe Song. 2013. "Consistent Classification of Image Time Series with Automatic Adaptive Signature Generalization." *Remote Sensing of Environment* 134 (July): 333–41. doi:10.1016/j.rse.2013.03.022.
- Greenfield, Eric J., David J. Nowak, and Jeffrey T. Walton. 2009. "Assessment of 2001 NLCD Percent Tree and Impervious Cover Estimates." *Photogrammetric Engineering & Remote Sensing* 75 (11): 1279–1286.
- Hargrove, William W., and Forrest M. Hoffman. 2004. "Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions." *Environmental Management* 34 (S1): S39–60. doi:10.1007/s00267-003-1084-0.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York. <http://link.springer.com/10.1007/978-0-387-84858-7>.
- Homer, Collin, Jon Dewitz, Joyce Fry, Michael Coan, Nazmul Hossain, Charles Larson, Nate Herold, et al. 2007. "Completion of the 2001 National Land Cover Database for the Conterminous United States." *Photogrammetric Engineering and Remote Sensing* 73 (4): 337.

- Homer, Collin, Chengquan Huang, Limin Yang, Bruce Wylie, and Michael Coan. 2004. "Development of a 2001 National Land-Cover Database for the United States." *Photogrammetric Engineering & Remote Sensing* 70 (7): 829–840.
- Huang, C., L. S. Davis, and J. R. G. Townshend. 2002. "An Assessment of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 23 (4): 725–49. doi:10.1080/01431160110040323.
- Huang, Chengquan, Limin Yang, Bruce K. Wylie, and Collin Homer. 2001. "A Strategy for Estimating Tree Canopy Density Using Landsat 7 ETM+ and High Resolution Images over Large Areas." <http://digitalcommons.unl.edu/usgspubs/111/>.
- Jepson, Wendy. 2005. "A Disappearing Biome? Reconsidering Land-Cover Change in the Brazilian Savanna." *Geographical Journal* 171 (2): 99–111. doi:10.1111/j.1475-4959.2005.00153.x.
- Johnson, Roger W. 2001. "An Introduction to the Bootstrap." *Teaching Statistics* 23 (2): 49.
- Kennedy, Robert E., Zhiqiang Yang, Justin Braaten, Catharine Copass, Natalya Antonova, Chris Jordan, and Peder Nelson. 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 (September): 271–85. doi:10.1016/j.rse.2015.05.005.
- Kennedy, Robert E., Zhiqiang Yang, Warren B. Cohen, Eric Pfaff, Justin Braaten, and Peder Nelson. 2012. "Spatial and Temporal Patterns of Forest Disturbance and Regrowth within the Area of the Northwest Forest Plan." *Remote Sensing of Environment* 122 (July): 117–33. doi:10.1016/j.rse.2011.09.024.
- Kim, Do-Hyung, Joseph O. Sexton, Praveen Noojipady, Chengquan Huang, Anupam Anand, Saurabh Channan, Min Feng, and John R. Townshend. 2014. "Global, Landsat-Based Forest-Cover Change from 1990 to 2000." *Remote Sensing of Environment* 155 (December): 178–93. doi:10.1016/j.rse.2014.08.017.
- Liu, Canran, Paul Frazier, and Lalit Kumar. 2007. "Comparative Assessment of the Measures of Thematic Classification Accuracy." *Remote Sensing of Environment* 107 (4): 606–16. doi:10.1016/j.rse.2006.10.010.
- Lunetta, Ross S, David M Johnson, John G Lyon, and Jill Crotnell. 2004. "Impacts of Imagery Temporal Frequency on Land-Cover Change Detection Monitoring." *Remote Sensing of Environment* 89 (4): 444–54. doi:10.1016/j.rse.2003.10.022.
- MacEachren, Alan M. 1992. "Visualizing Uncertain Information." *Cartographic Perspectives*, no. 13: 10–19.

- Martinuzzi, Sebastia'n. 2009. "Incorporating Remotely Sensed Tree Canopy Cover Data into Broad Scale Assessments of Wildlife Habitat Distribution and Conservation." *Journal of Applied Remote Sensing* 3 (1): 033568. doi:10.1117/1.3279080.
- Masek, Jeffrey G., Chengquan Huang, Robert Wolfe, Warren Cohen, Forrest Hall, Jonathan Kutler, and Peder Nelson. 2008. "North American Forest Disturbance Mapped from a Decadal Landsat Record." *Remote Sensing of Environment* 112 (6): 2914–26. doi:10.1016/j.rse.2008.02.010.
- Miller, Crane S., and Richard S. Hyslop. 1983. *California, the Geography of Diversity*. 1st ed. Palo Alto, Calif: Mayfield Pub. Co.
- Nemani, Ramakrishna R., and Steven W. Running. 1995. "Satellite Monitoring of Global Land Cover Changes and Their Impact on Climate." *Climatic Change* 31 (2–4): 395–413. doi:10.1007/BF01095154.
- Nowak, David J., and Eric J. Greenfield. 2010. "Evaluating The National Land Cover Database Tree Canopy and Impervious Cover Estimates Across the Conterminous United States: A Comparison with Photo-Interpreted Estimates." *Environmental Management* 46 (3): 378–90. doi:10.1007/s00267-010-9536-9.
- Olofsson, Pontus, Giles M. Foody, Stephen V. Stehman, and Curtis E. Woodcock. 2013. "Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty Using Stratified Estimation." *Remote Sensing of Environment* 129 (February): 122–31. doi:10.1016/j.rse.2012.10.031.
- Opitz, David, and Richard Maclin. 1999. "Popular Ensemble Methods: An Empirical Study." *Journal of Artificial Intelligence Research*, 169–198.
- Pidgeon, Anna M., Volker C. Radeloff, Curt H. Flather, Christopher A. Lepczyk, Murray K. Clayton, Todd Jerome Hawbaker, and Roger B. Hammer. 2007. "Associations of Forest Bird Species Richness with Housing and Landscape Patterns across the USA." *Ecological Applications* 17 (7): 1989–2010.
- Pontius, Robert Gilmore, and Marco Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing* 32 (15): 4407–29. doi:10.1080/01431161.2011.552923.
- Riemann, Rachel, Barry Tyler Wilson, Andrew Lister, and Sarah Parks. 2010. "An Effective Assessment Protocol for Continuous Geospatial Datasets of Forest Characteristics Using USFS Forest Inventory and Analysis (FIA) Data." *Remote Sensing of Environment* 114 (10): 2337–52. doi:10.1016/j.rse.2010.05.010.
- Sanyal, J, Song Zhang, J Dyer, A Mercer, P Amburn, and R J Moorhead. 2010. "Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty." *IEEE*

Transactions on Visualization and Computer Graphics 16 (6): 1421–30.
doi:10.1109/TVCG.2010.181.

- Schimel, David S., Vemap Participants, and B. H. Braswell. 1997. “Continental Scale Variability in Ecosystem Processes: Models, Data, and the Role of Disturbance.” *Ecological Monographs* 67 (2): 251–71. doi:10.1890/0012-9615(1997)067[0251:CSVIEP]2.0.CO;2.
- Shao, Yang, and Ross S. Lunetta. 2012. “Comparison of Support Vector Machine, Neural Network, and CART Algorithms for the Land-Cover Classification Using Limited Training Data Points.” *ISPRS Journal of Photogrammetry and Remote Sensing* 70 (June): 78–87. doi:10.1016/j.isprsjprs.2012.04.001.
- Sheeder, Scott A., Jeremy D. Ross, and Toby N. Carlson. 2002. “Dual Urban and Rural Hydrograph Signals in Three Small Watersheds1.” *JAWRA Journal of the American Water Resources Association* 38 (4): 1027–40. doi:10.1111/j.1752-1688.2002.tb05543.x.
- Strahler, Alan H., Luigi Boschetti, Giles M. Foody, Mark A. Friedl, Matthew C. Hansen, Martin Herold, Philippe Mayaux, Jeffrey T. Morisette, Stephen V. Stehman, and Curtis E. Woodcock. 2006. “Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps.” *European Communities, Luxembourg* 51. http://cndwebzine.hcp.ma/cnd_sii/IMG/pdf/Document2222222222-17.pdf.
- Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. “Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution.” *BMC Bioinformatics* 8: 25. doi:10.1186/1471-2105-8-25.
- Suriya, S., and B. V. Mudgal. 2012. “Impact of Urbanization on Flooding: The Thirusoolam Sub Watershed – A Case Study.” *Journal of Hydrology*, Hydrology Conference 2010, 412–413 (January): 210–19. doi:10.1016/j.jhydrol.2011.05.008.
- Theobald, David M., Scott J. Goetz, John B. Norman, and Patrick Jantz. 2009. “Watersheds at Risk to Increased Impervious Surface Cover in the Conterminous United States.” *Journal of Hydrologic Engineering* 14 (4): 362–68. doi:10.1061/(ASCE)1084-0699(2009)14:4(362).
- Townshend, John R., Jeffrey G. Masek, Chengquan Huang, Eric F. Vermote, Feng Gao, Saurabh Channan, Joseph O. Sexton, et al. 2012. “Global Characterization and Monitoring of Forest Cover Using Landsat Data: Opportunities and Challenges.” *International Journal of Digital Earth* 5 (5): 373–97. doi:10.1080/17538947.2012.713190.
- Turner, D. P., W. D. Ritts, B. E. Law, W. B. Cohen, Z. Yang, T. Hudiburg, J. L. Campbell, and M. Duane. 2007. “Scaling Net Ecosystem Production and Net Biome Production over a Heterogeneous Region in the Western United States.” *Biogeosciences Discussions* 4 (2): 1093–1135.
- Wang, Jie, Yuanyuan Zhao, Congcong Li, Le Yu, Desheng Liu, and Peng Gong. 2015. “Mapping Global Land Cover in 2001 and 2010 with Spatial-Temporal Consistency at 250m

- Resolution.” *ISPRS Journal of Photogrammetry and Remote Sensing* 103 (May): 38–47. doi:10.1016/j.isprsjprs.2014.03.007.
- White, J. C., M. A. Wulder, G. W. Hobart, J. E. Luther, T. Hermosilla, P. Griffiths, N. C. Coops, et al. 2014. “Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science.” *Canadian Journal of Remote Sensing* 40 (3): 192–212. doi:10.1080/07038992.2014.945827.
- Wickham, J.D., T.G. Wade, and D.J. Norton. 2014. “Spatial Patterns of Watershed Impervious Cover Relative to Stream Location.” *Ecological Indicators* 40 (May): 109–16. doi:10.1016/j.ecolind.2014.01.013.
- Wilson, B. Tyler, Andrew J. Lister, and Rachel I. Riemann. 2012. “A Nearest-Neighbor Imputation Approach to Mapping Tree Species over Large Areas Using Forest Inventory Plots and Moderate Resolution Raster Data.” *Forest Ecology and Management* 271 (May): 182–98. doi:10.1016/j.foreco.2012.02.002.

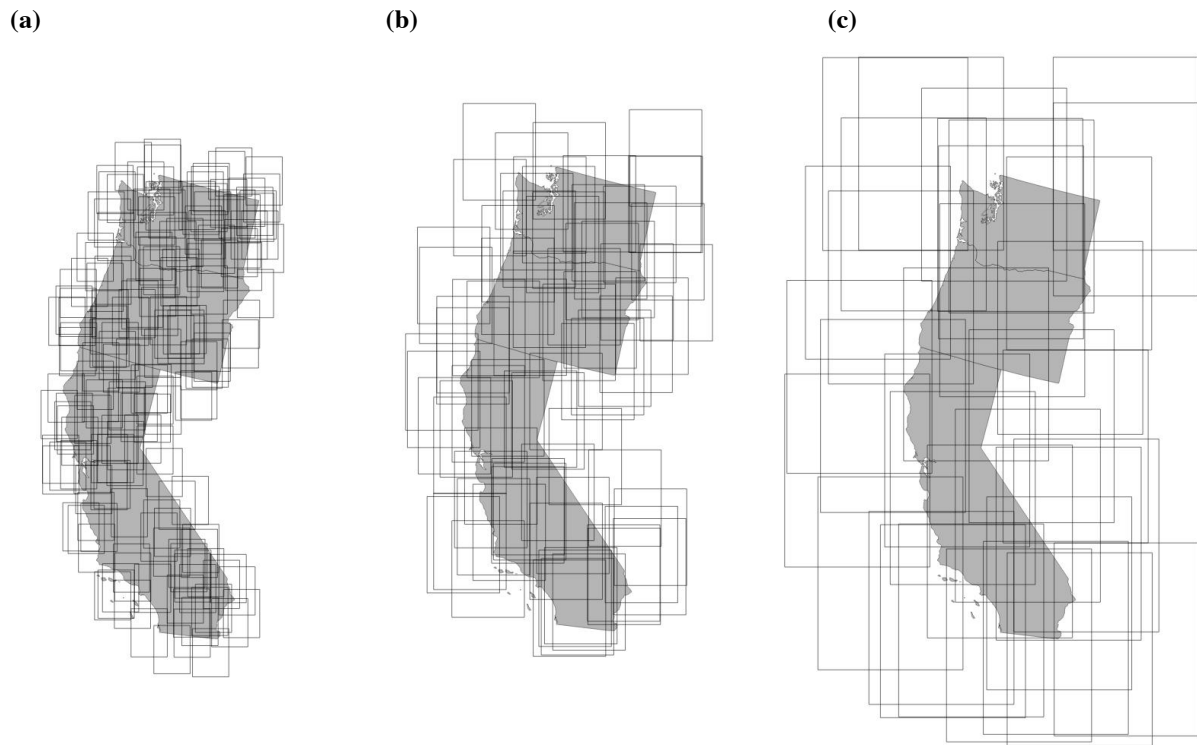
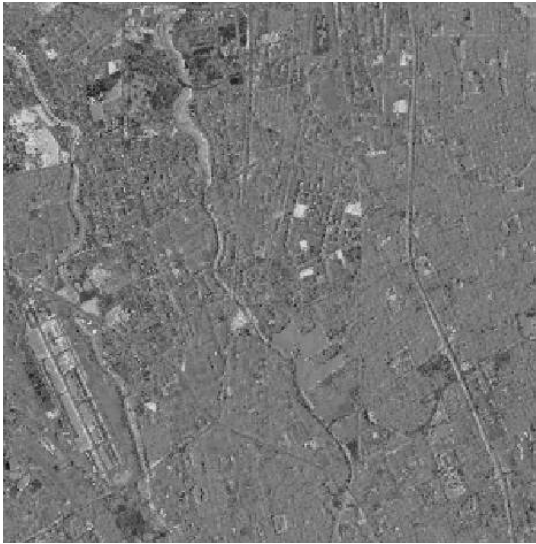
Figures

Figure 2.1. Examples of small (a), medium (b), and large (c) support sets. Support sets must be large enough to provide adequate information to estimators but also small enough to limit the complexity of the per-estimator data space. Areas of support sets outside the modeling region are ignored when training the model and predicting.

(a)



(b)

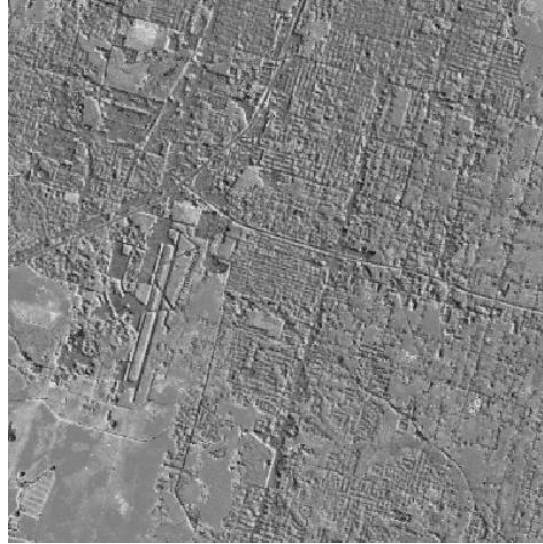


Figure 2.2. Difference images between STEM and NLCD maps in San Francisco, CA (a) and Seattle, WA (b). The embossed pattern is indicative of pixel misalignment in the source LandTrendr data. Pixel offsets differed through the modeling region, so this error could not be directly corrected in the source data.

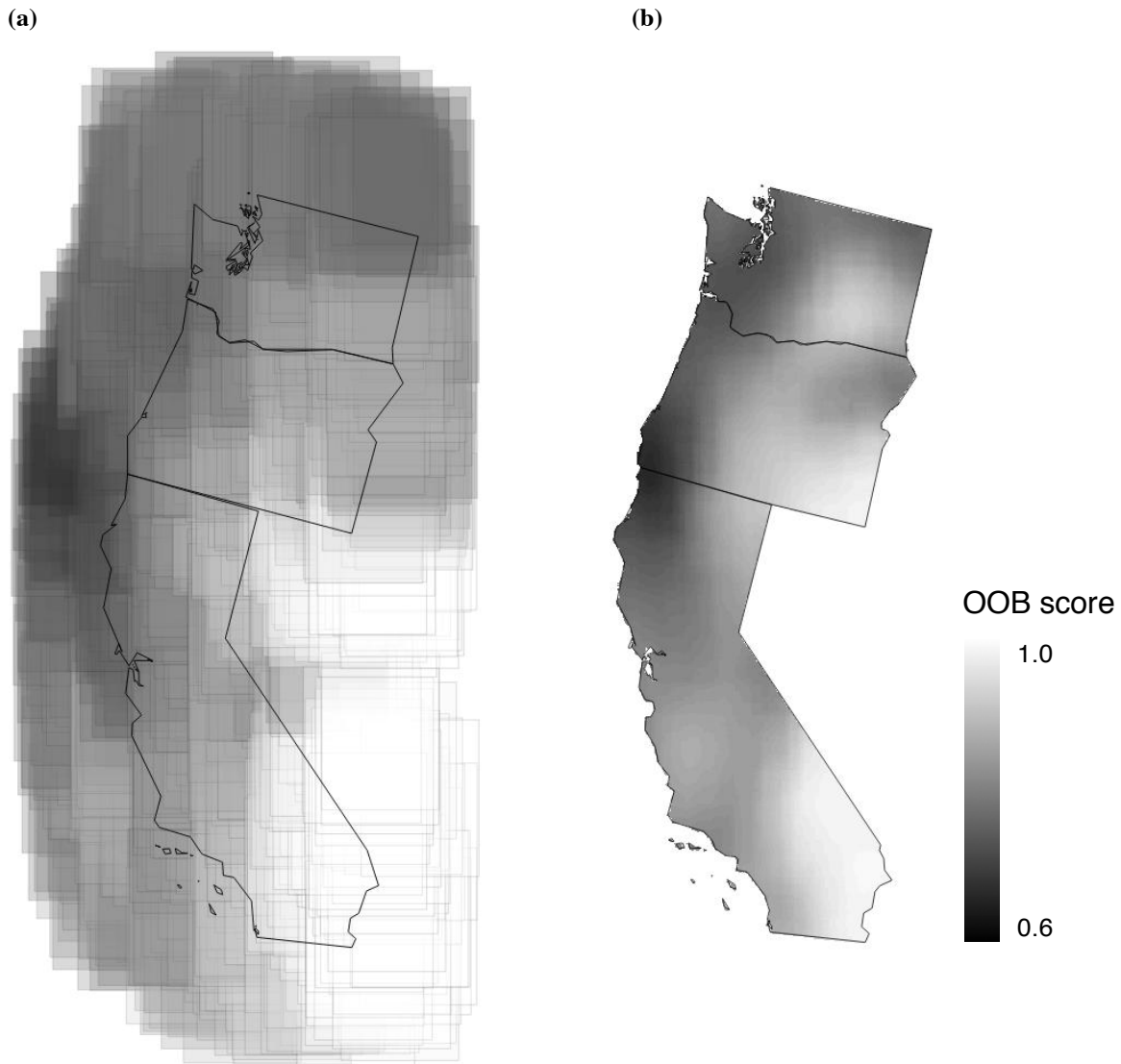


Figure 2.3. Illustration of computing an OOB map from a STEM. All pixels of each support set are first assigned the value of its OOB accuracy rate (a). Each pixel in the final OOB map (b) is the average of all overlapping pixels from support sets. Although many support sets extend beyond the modeling region, only data within the modeling region are used to construct the model and, thus, calculate the OOB accuracy rate for each estimator.

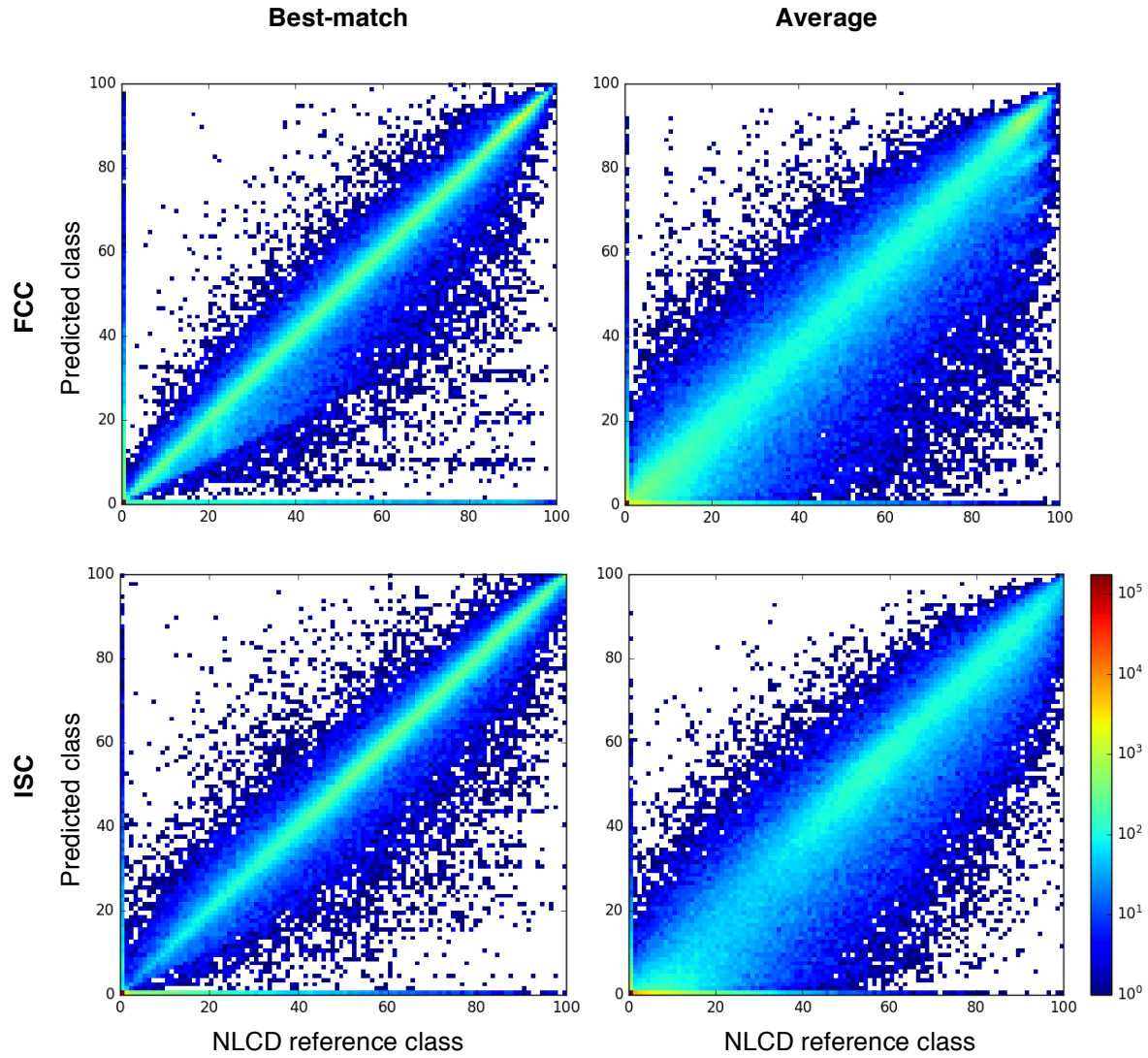


Figure 2.4. Two-dimensional histograms comparing STEM predictions of FCC (top row) and ISC (bottom row) to NLCD reference samples using the best-match method (left column) and average method (right column). The narrow concentrated plots of the best-match method showed better agreement with NLCD reference data for both FCC and ISC.

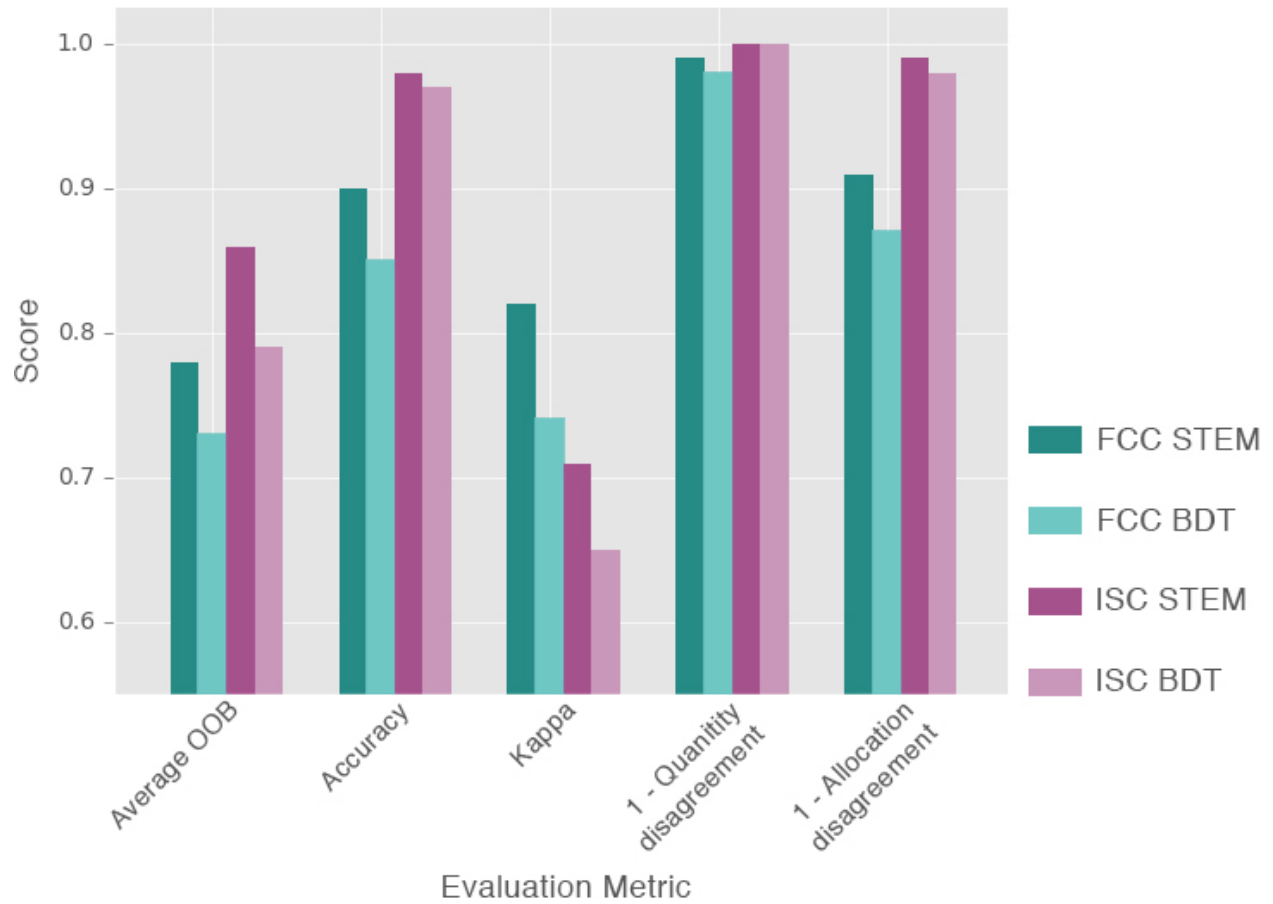


Figure 2.5. All five summary statistics for the FCC and ISC BDTs and the best performing STEMs. Quantity and allocation disagreement are both subtracted from 1 to make their scales comparable to other metrics. STEMs outperformed BDTs according to every evaluation metric.

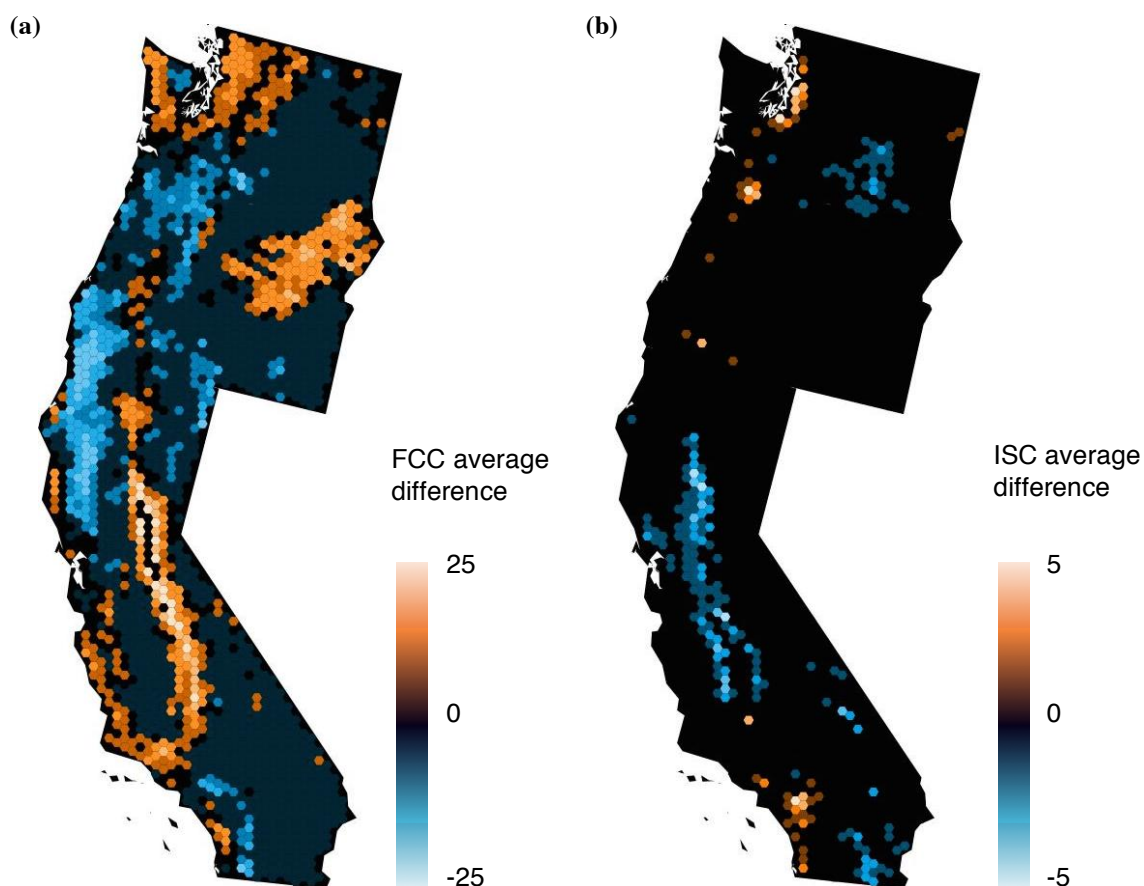


Figure 2.6. Difference of STEM and BDT maps averaged by 35,000 ha hexagons. Negative values occur where the BDT prediction was higher, and positive values indicate that the STEM prediction was higher. The FCC BDT (a) generally over-predicted in dry forests and under-predict in wet forests. The ISC BDT (b) under-predicted near some urban centers and over-predicted in some agricultural areas.

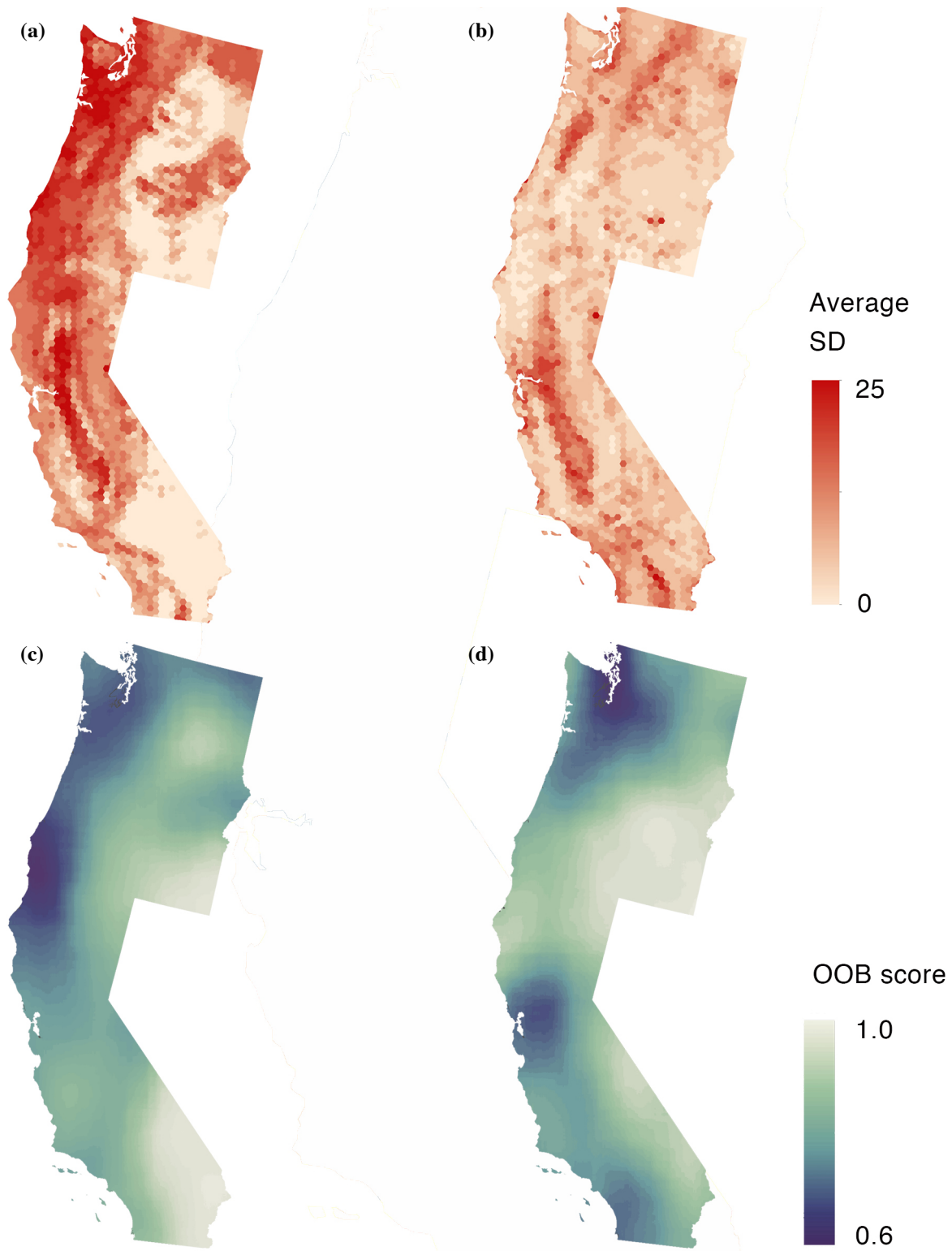


Figure 2.7. Standard deviation (top row) and OOB score maps (bottom row) for FCC (left column) and ISC (right column) models. Standard deviation is shown here averaged by 35,000 ha hexagons so it only ranges from 0-25%.

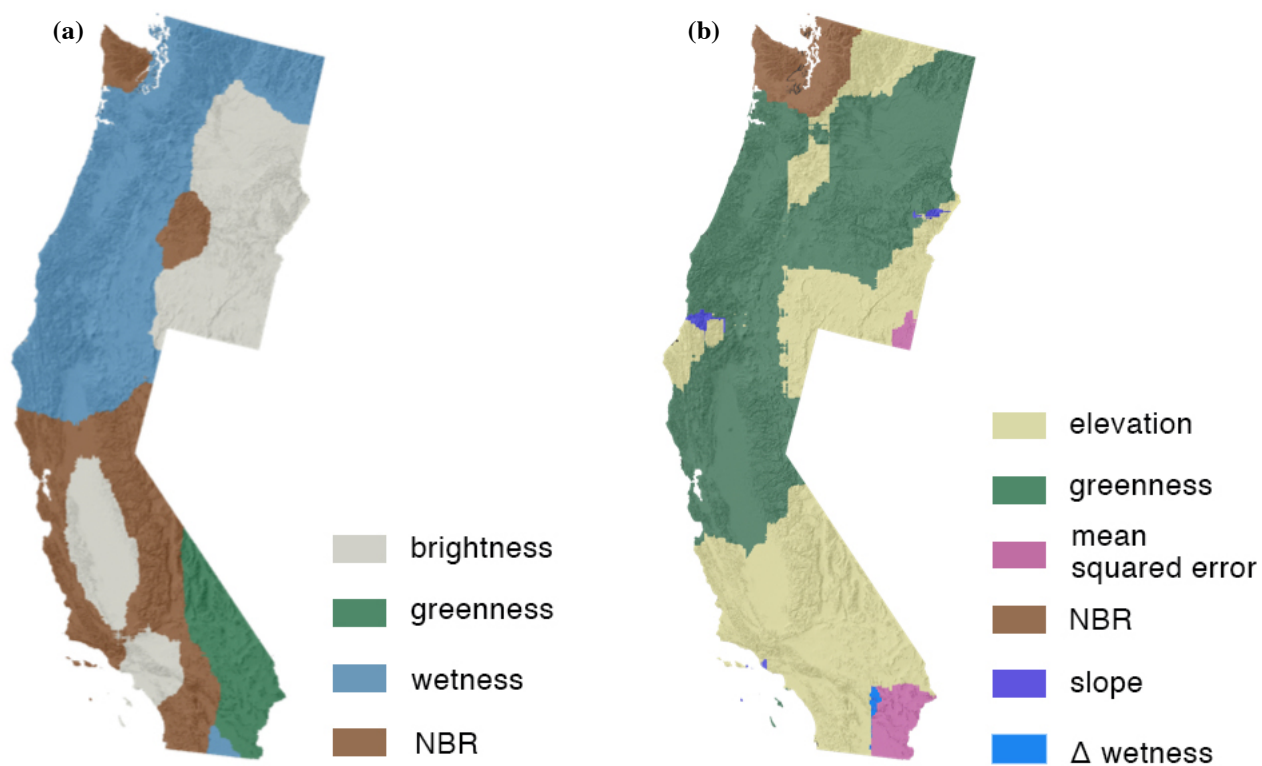


Figure 2.8. Variable importance maps for FCC (a) and ISC (b). We derived variable importance per pixel to assess how predictor-response relationships changed across the modeling region.

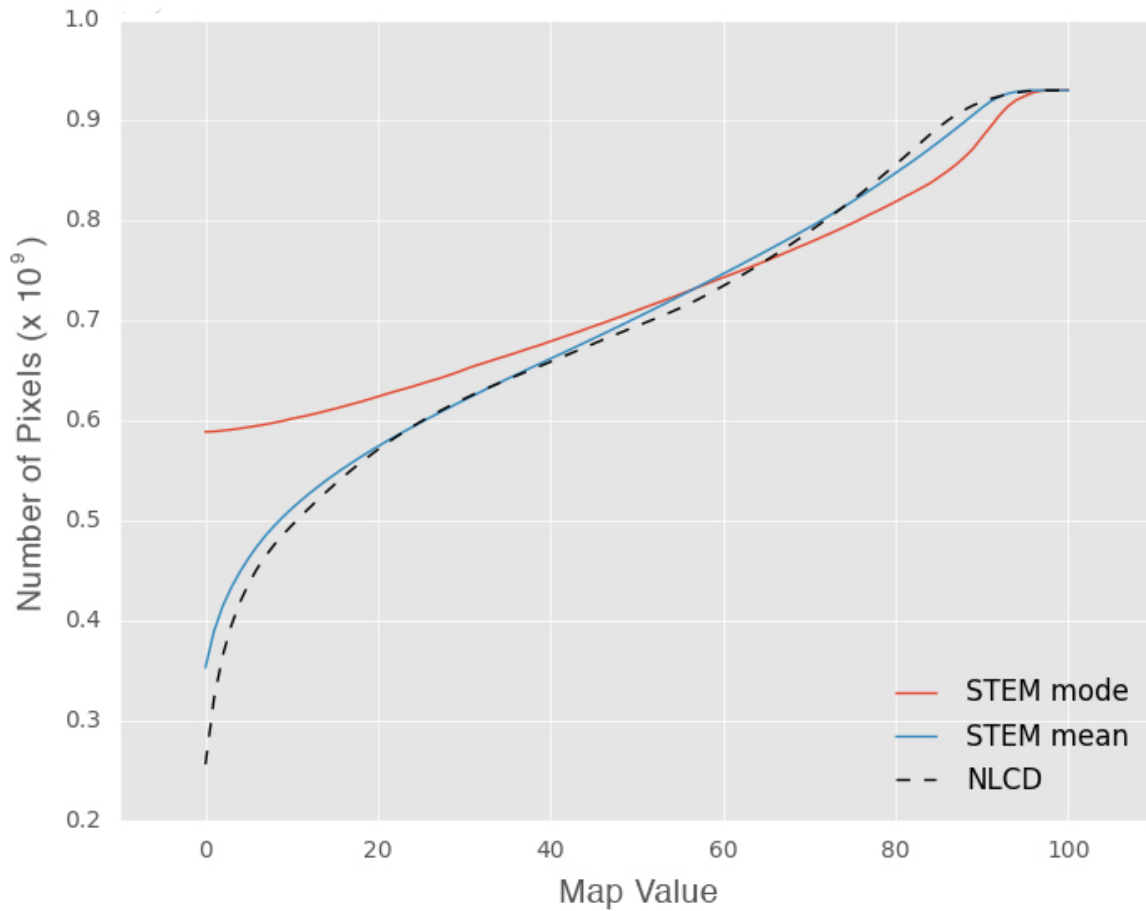


Figure 2.9. Cumulative distribution functions for 2011 FCC STEM mode, STEM mean, and NLCD. The 2011 STEM mean prediction agreed much more with 2011 NLCD than the STEM mode. We found that that the 2001 STEM mode, however, agreed much more with 2001 NLCD than STEM mean.

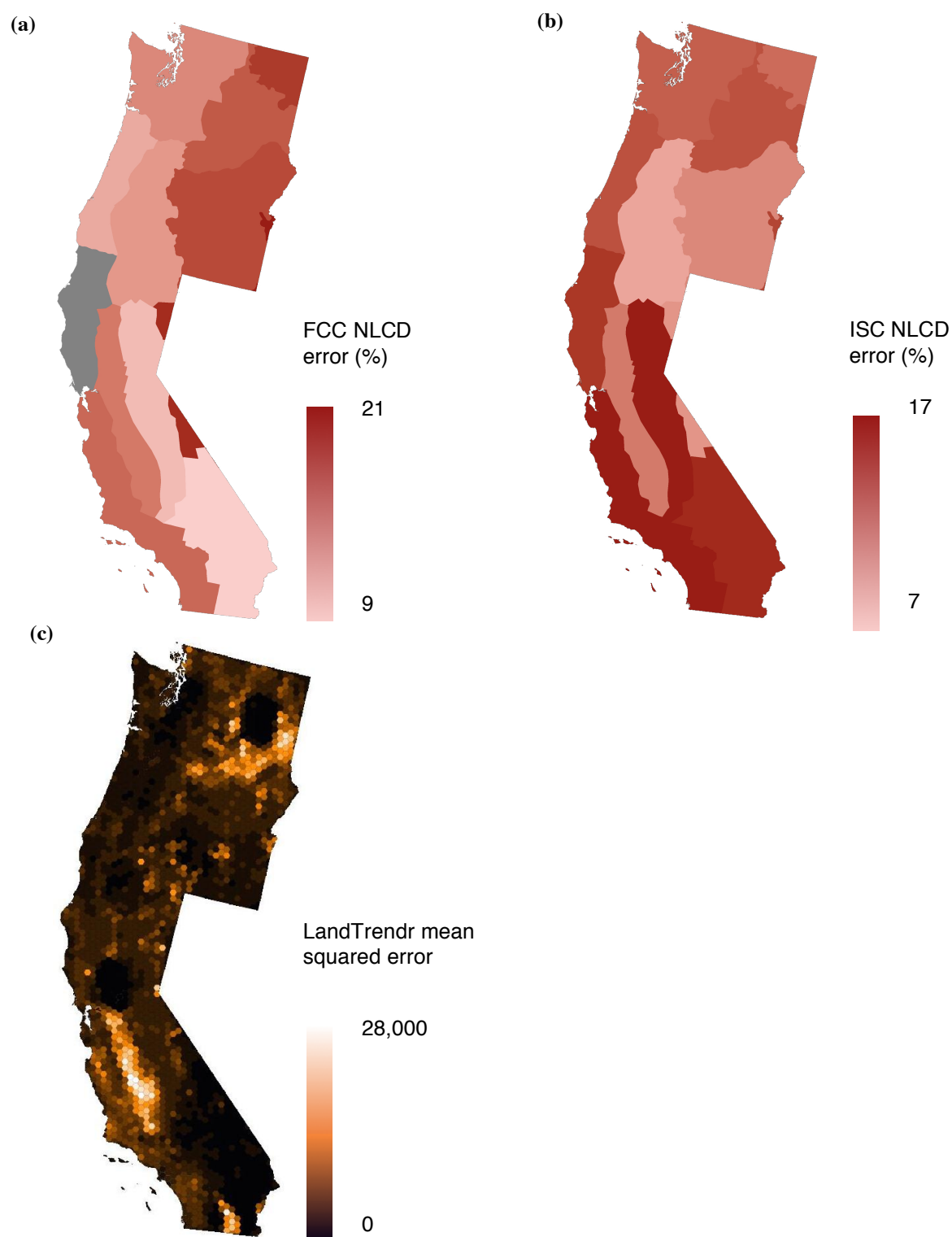


Figure 2.10. Error in source data: (a) FCC NLCD cross-validation error, (b) ISC NLCD cross-validation error, and (c) 2001 LandTrendr mean squared error. NLCD error for FCC was not reported for the modeling region colored gray (Homer et al. 2007). LandTrendr mean squared error maps are in units of squared rescaled spectral values.

Tables

Table 2.1. Area-estimated confusion matrix for the best performing FCC model. All cells in the matrix are represented as a percent of the total area of the reference map. NLCD reference classes are columns and STEM predicted classes are rows. Overall accuracy for this model was 89.6% and the kappa coefficient was 0.82.

		Reference map class										Total ref.	User's (%)	Area (ha)
		0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100			
Predicted map class	0–10	0.619	0.004	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0	0.634	97.6	53724446
	11–20	0.008	0.014	0.006	0.001	0	0	0	0	0	0	0.030	46.8	2012982
	21–30	0.002	0.001	0.019	0.006	0.002	0	0	0	0	0	0.031	63.0	2422968
	31–40	0	0	0.002	0.021	0.005	0.001	0	0	0	0	0.032	66.9	2584567
	41–50	0	0	0	0.003	0.024	0.005	0	0	0	0	0.034	69.4	2835910
	51–60	0	0	0	0	0.003	0.027	0.005	0	0	0	0.037	71.6	3043708
	61–70	0	0	0	0	0	0.003	0.03	0.006	0	0	0.040	73.9	3270071
	71–80	0	0	0	0	0	0	0.003	0.037	0.006	0	0.046	79.4	3851839
	81–90	0	0	0	0	0	0	0	0.003	0.064	0.004	0.071	89.5	6061048
	91–100	0	0	0	0	0	0	0	0	0.003	0.043	0.046	93.3	3868383
Total pred.		0.631	0.019	0.03	0.033	0.036	0.038	0.041	0.049	0.076	0.049			
Producer's (%)		98.2	73.5	63.3	63.8	66.4	69.9	72.6	75.9	83.6	88.5			

Table 2.2. Area-estimated confusion matrix for the best performing ISC model. All cells in the matrix are represented as a percent of the total area of the reference map. NLCD reference classes are columns and STEM predicted classes are rows. Overall accuracy for this model was 98.0% and the kappa coefficient was 0.71.

		Reference map class										Total ref.	User's (%)	Area (ha)
		0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100			
Predicted map class	0–10	0.9420	0.0022	0.0006	0.0003	0.0002	0	0	0	0	0	0.945	99.6	811502308
	11–20	0.0048	0.0031	0.0006	0	0	0	0	0	0	0	0.009	35.9	4333119
	21–30	0.0018	0.0008	0.0030	0.0006	0	0	0	0	0	0	0.006	48.0	4042754
	31–40	0.0008	0	0.0007	0.0028	0.0006	0	0	0	0	0	0.005	54.8	3676774
	41–50	0.0004	0	0	0.0005	0.0031	0.0005	0	0	0	0	0.005	66.0	3726687
	51–60	0.0002	0	0	0	0.0004	0.0031	0.0004	0	0	0	0.004	72.0	3483121
	61–70	0.0002	0	0	0	0	0.0004	0.0023	0.0003	0	0	0.003	72.5	2538905
	71–80	0	0	0	0	0	0	0.0003	0.0015	0.0002	0	0.002	73.0	1645476
	81–90	0	0	0	0	0	0	0	0.0001	0.0011	0	0.001	78.3	1135276
	91–100	0	0	0	0	0	0	0	0	0	0.0006	0.001	85.0	522241
Total pred.		0.950	0.006	0.005	0.004	0.005	0.004	0.003	0.002	0.001	0.001			
Producer's (%)		99.1	49.9	59.1	64.6	69.4	73.4	73.7	75.5	80.3	85.0			

Table 2.3. Summary statistics for all FCC and ISC models. We found decreasing accuracy with smaller sample size and increasing support set size.

Number of samples	Set size (km)	OOB score	Overall accuracy	Kappa	Quantity disagreement	Allocation disagreement
FCC						
100,000	400 x 300	0.733	87.4	0.792	0.018	0.108
300,000	400 x 300	0.745	87.6	0.800	0.018	0.105
600,000	400 x 300	0.753	88.0	0.802	0.016	0.103
1,000,000	400 x 300	0.757	88.1	0.803	0.017	0.095
1,500,000	400 x 300	0.782	89.0	0.810	0.017	0.093
1,500,000*	200 x 150	0.803	89.8	0.822	0.014	0.090
1,500,000	800 x 600	0.729	88.2	0.794	0.019	0.010
ISC						
100,000	400 x 300	0.733	95.7	0.630	0.019	0.022
300,000	400 x 300	0.753	96.3	0.655	0.018	0.018
600,000	400 x 300	0.762	96.1	0.629	0.020	0.019
1,000,000	400 x 300	0.770	96.2	0.629	0.019	0.018
1,500,000	400 x 300	0.840	98.1	0.710	0.003	0.016
1,500,000*	200 x 150	0.855	98.0	0.713	0.005	0.014
1,500,000	800 x 600	0.820	98.1	0.706	0.002	0.016

*Sample and support set size selected to create time series

CHAPTER 3: INFLUENCES OF TEMPORAL RESOLUTION ON AVIAN SPECIES DISTRIBUTION MODELS

Introduction

Understanding the distribution of a species is critical for effective conservation planning and land management. To mitigate the effects of habitat degradation, climate change, and other deleterious impacts, conservationists and land managers must know how and why species distributions change over time. Species distribution models (SDMs) are one of the most widely used tools for such habitat analysis. These models must be applicable across broad extents of a species' range in space and time to be effective for conservation and management (Betts et al. 2006; Schröder and Richter 2000).

As with many issues in landscape ecology, development of an SDM requires careful consideration of spatial and temporal scale. Improper selection of scale can lead to erroneous species-environment relationships and a poorly specified SDM (Mitchel et al. 2001). Many studies have demonstrated that the spatial scale used to develop an SDM can significantly influence the outcome of a model (e.g., Austin and Van Neil 2011; Ferraz et al. 2012; Franklin and Miller 2009, 59; Gottschalk et al. 2011; Mitchel et al. 2001; Scott 2002; Seo et al. 2009; Song et al. 2013). Differences in scale for an SDM are often expressed as the resolution of response and predictor data used to develop the model. Moreover, SDMs must be produced at a scale relevant to their application, and many authors have argued that the spatial resolution of predictor data must match spatial scales important to the species considered (Addicott et al. 1987; Betts et al. 2006; Mitchel et al. 2001).

Although spatial resolution of SDMs has received significant attention, few studies have offered temporal resolution the same consideration. In the present study, we express temporal resolution as the time interval between representative dates of predictor or response data. Many studies rely on field data from wide temporal extents, often including data from multiple years. Rarely, however, are field data from different years accompanied by concomitant predictor data of a comparable temporal resolution or extent (e.g., Fink et al. 2010; Gschweng et al. 2012; Gutzwiller and Barrow 2001; Rullman and Marzluff 2014). Indeed, the influence of differences in temporal resolution and extent between predictor and response data remains largely unexamined (Elith and Leathwick 2009, 689; Radchuk et al. 2014). Ives and Klopfer (1997) and Adler and Lauenroth (2003) argue that understanding spatial patterns of species distributions often requires understanding temporal patterns.

Nevertheless, a relatively small number of studies have investigated the temporal dynamics of SDMs. Bulluck and others (2006) found that temporal variation of SDMs of breeding birds in the Great Basin had greater influence on model outcomes than spatial variation. Consequently, they concluded that temporal patterns of distribution were more varied than spatial patterns. Other studies have also documented high year-to-year variability in species occurrences in desert ecosystems. For instance, Gutzwiller and Barrow (2001) found dramatic inter-annual differences in bird distributions in the Sonoran Desert. Much of this variation was due to yearly changes in environmental conditions. Gschweng and others (2012) determined that a model of Elanor's falcon winter distributions in Madagascar produced with a normalized difference vegetation index (NDVI) at a monthly resolution predicted better than a model with NDVI for the whole

winter. Gschwend and others note that Eleonara's falcons and other species may respond especially strongly to phenological changes, and a more temporally coarse model may be suitable for species with distributions more closely related to variables such as land cover, which show less intra-annual variation. Vierling and others (2014) developed avian SDMs trained on predictors that were contemporaneous with field observations and compared them to SDMs trained on predictors with a six-year lag. They found a slight decrease in model accuracy with predictors representing vegetation structure six years before the response data were collected. In contrast, Sheeren, Bonthoux, and Balent (2014) found that SDMs produced with NDVI derived from an image with an acquisition date that most accurately reflects vegetation structure, rather than an image contemporaneous with the date of bird observation, produced the best model. This previous body of research clearly demonstrates a strong relationship between the representation of temporal dynamics in an SDM and model performance.

Despite limited investigation of the influence of temporal dynamics of predictors in SDMs, these models often rely on land cover and other spatially explicit data that are highly temporally dynamic. Indeed, a well-established relationship exists between species prevalence and land use/land cover (Manel et al. 2000; Mitchel et al. 2001). Rullman and Marzluff (2014) found that SDMs based only on land cover performed better for some species than those that included other predictors like prey abundance. Models examined by that study that excluded land cover consistently performed lowest for all species modeled. Moreover, land cover changes frequently; the innumerable studies in the field of land change detection are a testament to this (e.g., Cohen et al. 2010; Homer et al. 2015; Huang et al. 2010; Kennedy et al. 2012, Kim et al. 2014; Masek et al. 2008; Pielke et al. 2011; Wang et al. 2015, Xin et al. 2013). A comparison of two years of

data from the National Land Cover Database also demonstrates this highly dynamic quality (Figure 3.1). For SDMs, inaccurate training data can lead to erroneous estimations of species-environment relationships (Vaughan and Ormerod 2003), and variable environmental conditions can cause significant changes in species distributions (Gutzwiller and Barrow 2001; Hejl and Beedy 1986; Johnson 1995). Additionally, variability in predictor data is likely to increase with the spatial and temporal extents of a modeling region (Fink et al. 2010).

A likely reason for the small number of studies investigating temporal resolution is the lack of available predictor data at fine temporal resolutions. Much of these data are products of remotely sensed phenomena, and in developing remote sensing instruments, designers usually must address a tradeoff between spatial resolution or extent and temporal resolution (Wang et al. 2015; Xin et al. 2013). This, in combination with previously unavailable computer processing capabilities, has made broad scale data at both fine spatial and temporal resolutions rare (Zhu and Li 2013). For example, the most commonly used land cover data in the United States, the National Land Cover Database, is only produced in full every five to eight years. Recent developments in digital computational efficiency and the public release of the entire Landsat archive have made possible many broad-scale, fine-resolution mapping initiatives (e.g., Cohen and Goward 2004; Gray and Song 2013; Huang et al. 2010; Kennedy et al. 2012; Kim et al. 2014; Townshend et al. 2012).

Our study leveraged the fine spatial and temporal resolution of two novel datasets to assess the influence of temporal resolution of predictor data on avian SDMs. We elected to model distributions of birds to specifically take advantage of the unique combination of broad extent

and fine resolution observations from eBird, a citizen science project comprising an online database of birding checklists submitted by users worldwide (Dickinson et al. 2010). Predictor data were mostly derived from LandTrendr, an approach to mapping yearly land cover and land cover change (Kennedy et al. 2015). By adjusting the time interval between predictor datasets but still training models with yearly bird response data, we evaluated how the disparity between the two can affect model performance. We anticipated that response to changes in temporal resolution would vary for different species, but overall, finer temporal resolutions that represented environmental conditions contemporaneous with observations would produce more accurate models. We anticipate this work will shed new light on species-environment relationships, and provide insight on appropriate temporal resolutions for SDMs and sample design of bird monitoring programs.

Methods

Study area

Our study area covered all of California, Oregon, and Washington on the west coast of the conterminous United States. Both Oregon and Washington are broadly characterized by two vegetation zones, forest and steppe (Franklin and Dyrness 1988), generally corresponding to maritime and continental climatic zones, respectively. Vegetation zones are similarly divided in California by the southern Cascades and Sierra mountains, although more xeric vegetation communities are common in southern California with coastal chaparral in the west (Ashbaugh 1994, 110) and desert scrublands throughout much of southeastern California (Miller and Hyslop 1983). Elevations across the study area range from sea level to 4400 m. Agricultural land dominates much of the low-elevation central valleys throughout the study area with a wide

variety of crop types. Most urban development throughout Oregon and Washington is relegated to major urban centers and highway travel corridors (Ashbough 1994, 382). The same general pattern is found in CA, although development intensity and extent are generally greater across the state.

Bird observation data

The response data we used in this study came from the eBird reference dataset (ERD), a subset of the eBird citizen science project (Munson et al. 2013). Each record of the ERD is a checklist from a separate sampling event uploaded by one or more volunteers who reported counting all species present at the time of observation. In this way, the ERD represents presence-absence data, as all species not detected are presumed absent during the sampling event. eBird data are unique in North America in both their fine resolution and broad extent—both spatial and temporal (Hochachka and Fink 2012). Leveraging the power of crowd-sourced information, the ERD provides species occurrence data that would otherwise be impractical to collect.

Our primary goal in this study was to test the effect of temporally coarse predictor data on the performance of SDMs. Because there are numerous stochastic factors that could influence an SDM, we attempted to simplify our models to limit the influence of anything that might affect performance other than annual temporal resolution. We therefore restricted observations to the month of June from 2002–2012. The breeding season represents a time of relative stability in species occurrence (Fink et al. 2010), so the influence of seasonal dynamics on SDM performance would be minimal. Land cover predictors (discussed below) were also developed using summer Landsat imagery, so land cover would also be contemporaneous with these bird

observations. We also used only observations reported by observers adhering to eBird's "point count" protocol whereby observers remain in a single location during the sampling event.

Despite the unprecedented scope and utility of eBird data, there is much uncertainty about the ability of non-experts to accurately identify bird species (Yu, Wong, and Kelling 2014).

Uncertainty in detection probability, in particular, can have prominent effects on SDM accuracy (McClintock et al. 2010). Multiple eBird participants frequently record observations at the same locations either at different times of the same day or different days within the same month. A potential result of this is that both positive (species was present) and negative (species was absent) observation co-occurred, potentially making predictor-response patterns indiscernible. To resolve this, we made the simplifying assumption that if a bird was observed at a location in a given year, it was suitable habitat for that species during that year. We removed any negative observations from any locations where a positive observation had also been reported in the same year. This likely had the consequence of flattening any diurnal patterns of detection (Johnson 2008), but since we were mostly concerned with adequately representing species habitat quality in our training data, this would not affect our results. These curatorial steps on the training data left 17,544 observations from 5,728 unique locations (Figure 3.2).

We modeled breeding season distributions for five migratory bird species: 1) White-crowned Sparrow (*Zonotrichia leucophrys*), 2) Barn Swallow (*Hirundo rustica*), 3) Hermit Thrush (*Catharus guttatus*), 4) Western Meadowlark (*Sturnella neglecta*), and 5) Black-headed Grosbeak (*Pheucticus melanocephalus*). All five species are easily identifiable birds with few, if any, co-occurring species similar in appearance or sound. We intentionally selected two species

associated with forests (Hermit Thrush and Black-headed Grosbeak) and two birds more commonly found in anthropogenic habitats that are more persistent from year to year during the temporal extent of the study, such as urban and agricultural environments (Barn Swallow and Western Meadowlark). We were interested in seeing if habitat preference was related to performance with differing temporal resolutions, as different land cover types exhibit varied rates of disturbance (Homer et al. 2015). White-crowned sparrows are commonly associated with urban and agricultural landscapes outside the breeding season, but during the breeding season isolated populations are common in disturbed forests of the conterminous United States. We selected this species as a more complex example of habitat preference. Additionally, actual occurrence of Black-headed Grosbeaks is anecdotally stochastic, with birds often unexpectedly absent in locations they inhabited in previous years. We, therefore, expected model performance for this species to deteriorate more with decreased temporal resolution.

Predictor data

Because we modeled distributions for multiple species, we used a set of predictors generic enough to represent ecological relationships with environmental conditions for all species considered. Land cover is often central to an SDM (Manel et al. 2000; Mitchel et al. 2001; Rullman and Marzluff 2014), as it is in some sense a coarse description of habitat. Our model relied heavily on various land cover metrics derived from products of LandTrendr, an approach for mapping yearly land cover information from Landsat time-series data (Kennedy et al. 2010). LandTrendr is unusual among land cover change detection algorithms in that it employs temporal segmentation, which allows for detection of both abrupt changes (i.e., short, fast disturbances such as wildfire) and gradual trends (i.e., long, slow disturbances such as insect

damage). The spectral trajectory of each pixel is idealized as a series of fitted straight-line segments, which omits noise while retaining essential detail. For every year in the time-series stack, fitted images are created by interpolating each pixel's value to the appropriate time step along these segments. Each pixel is ultimately fitted to its own trajectory (Kennedy et al. 2015). The end product is a smoothed time series of spectral images that can be used for developing other yearly land cover datasets.

The land cover maps we used as predictors for our models included canopy cover and impervious surface represented as a percent of coverage for each pixel (see Chapter 2). We also described habitat with classified land cover data separated into seven discrete classes: water, snow/ice, urban, bare ground, deciduous forest, coniferous forest, and shrub (unpublished LandTrendr data). Since a single scale is often insufficient to represent ecological preference of a species (Betts et al. 2006), we additionally evaluated the percent of cover within 1.2 ha (5-pixel diameter) and 13.5 ha (15-pixel diameter) circular neighborhoods for each land cover class and for forest canopy cover and imperviousness. Fink and others (2010) found these neighborhood sizes to be effective for a generic avian SDM.

In addition to yearly land cover predictors, we also used several variables for which we did not adjust the temporal resolution, meaning we represented each with a single map for all years. These included elevation from the National Elevation Dataset (Gesch et al. 2002), ecoregion boundaries (US Environmental Protection Agency 2012), human population density (US Census Bureau 2013), and three climatic variables from 30-year averages: mean annual precipitation, mean annual minimum temperature, and mean annual maximum temperature (PRISM 2012).

Although inter-annual climatic variation can dramatically impact spatiotemporal variation in bird occurrence (Gutzwiller and Barrow 2001), our intent in including climatic variables was to offer a more detailed description of habitat across the study area. For this purpose, 30-year averages more adequately described ecoregional differences in habitat types across such a broad extent. Similar to climatic variables, changes in population density could affect distributions of certain species. Population densities were calculated from census tracts, and the effective differences in inter-annual population densities are likely minimal at that spatial scale. The intent of including ecoregions in the model was to further describe spatial variation in species occurrence. Ecoregions offered a coarse description of geographic position while simultaneously reflecting potentially important changes in ecological gradients. All predictors are listed in Table 1.

Lastly, we included two predictors provided with each record in the ERD. To account for variation in detection rates we included the amount of time spent for each observation, as the detection probability increases with observation duration (Diefenbach et al. 2007). To explain inter-annual differences in occurrence, we included the year of the observation as a predictor in all models.

Model description

We developed predictive models of bird distributions using a random forest model (Breiman 2001). A random forest is an ensemble model, meaning the final prediction is an aggregation of a population of estimators. Random forests, like many ensemble methods, rely on randomization to approximate different states of a system (Araújo and New 2007). By averaging over an ensemble of slightly randomized estimators, a random forest can, in theory, achieve a final

prediction that is highly representative of the actual system. For modeling probabilities of species occurrence with binary response data, our final predictions ranged continuously from 0 to 1, with a value of 0 indicating the species was absent.

The base estimator algorithm of a random forest is a decision tree, an algorithm that builds a model by recursively splitting training samples and constructing rules for how to separate the samples at each split (Breiman 1984). The goal of all predictive models is to determine the patterns between predictor variables and a response variable, and use those patterns to interpolate for ‘out-of-sample’ locations. Each rule of a decision tree can be considered an estimation of the relationship between a given predictor and the response variable. The ‘best’ split is determined according to a given information criterion commonly referred to as impurity. At each split, a decision tree uses a single variable that produces the lowest impurity (*ibid*) (i.e., establishes the greatest discrimination between resulting classes of the split).

The random forest algorithm introduces randomness into each decision tree estimator in two ways: 1) each estimator is trained on a random subset of approximately 63% of all training samples (i.e, a bootstrap sample), and 2) at each split, a random subset of predictors is selected, and the ‘best’ split is determined from this subset (Breiman 2001). The process of training estimators on bootstrap samples and aggregating is known as “bagging” (*bootstrap-aggregating*) (Breiman 1996a). With a bagged model, each training subset likely excludes some outliers or corrupt samples. Thus, over multiple iterations the training data come to more closely represent the actual system (Breiman 1996b; Dietterich 2000). Additionally, the strength of an ensemble method lies in the diversity of the estimator population, and training each estimator on a different

set of samples introduces such diversity. Random selection of predictors at each split, also known as the random subspace method (Ho 1998), facilitates diverse estimators for similar reasons (Deitterich 2000). We developed and evaluated our models in the Python programming language and used the Scikit-learn implementation of the random forest regressor algorithm (Pedregosa et al. 2011).

A concern with modeling spatial distributions of any phenomenon across large extents is that the relationships between predictor and response variables are likely to change with the extent of a modeling region. Modeling approaches that are sensitive to local scale variation, then, may be more appropriate for broad-scale models (see Chapter 2). Fink and others (2010) demonstrated that an ensemble of locally defined estimators may be better suited for predicting intra-annual, broad-scale distributions for migratory birds, but only for modeling distributions for certain times of the year. For times when bird distributions were more stable (e.g., the breeding season), a model with globally defined estimators performed comparably to a locally sensitive model. Since we modeled June distributions, a random forest with globally defined estimators should perform equally well for our study.

Adjusting temporal resolution

The optimal temporal resolution of an SDM is likely dependent on several factors including the behavior, habitat preferences, and tolerance to disturbance of the species considered. For this reason, our study compared multiple temporal resolutions for each species of which we modeled occurrence probabilities. We expressed temporal resolution as the time interval between years of predictor data, but for all temporal resolutions, we used bird observations from all years. For

observations from a year that did not match a year of predictor data, we simply used the closest year for that resolution. For instance, if we used only predictors from 2002, 2007, and 2012 (representing a 5-year resolution) and an observation occurred in 2005, the predictor values for that training sample would be drawn from 2007 data. If an observation was equidistant from two years of predictor data, we used the earlier year. For all species we tested resolutions of 1, 2, 5, and 10 years.

Once we trained the models, we used only 2011 land cover maps to make predictions, effectively producing species occurrence maps representing 2011. This year is well represented in the ERD observations (see Figure 3.7), but only the 1-year resolution model used training data with predictors drawn from this year. We believed this would adequately demonstrate each model's ability to generalize for temporal patterns in response and predictor data.

Model evaluation

Effective model evaluation requires an independent set of samples not used to train the model. The intent is to assess how well a model can generalize for locations and predictor-response patterns not seen during model training. A common criticism of eBird data is that they are more likely to be collected near roads, trails, and other developments (Hochachka and Fink 2012). Also many of the same locations in the ERD are repeatedly sampled on different dates. To address these spatial biases, users are encouraged to visit locations with little data coverage (Boakes et al. 2010). Nevertheless, any evaluation of a model trained on and validated with these data must address the inherent spatial bias to ensure the entire modeling region is represented in the evaluation process. Our sampling scheme addressed these spatial biases in the ERD in two

ways. First, we split training and evaluation sets such that they did not share any common locations, reserving 20% of locations for evaluation and 80% for training. Second, Fink and others (2010) proposed a straightforward subsampling algorithm for model evaluation that relies on a geographically stratified design. First, a grid of a specified, uniform cell size is randomly located across the modeling region. From each grid cell, a specified number of samples are randomly drawn from the test set. These random samples are used to statistically evaluate the model. The process of sampling and evaluating is repeated for a given number of Monte Carlo iterations, averaging the statistical evaluation metric(s) across all iterations. We used a grid cell size of 25 x 18 km, a per-cell sample size of 2, and 200 Monte Carlo iterations. Geographically stratifying the subsampling scheme effectively offers more even coverage across the modeling region than other random sampling schemes might.

To compare results from models of differing temporal resolutions, we used only testing samples from 2011 observations, since we made predictions for 2011 occurrences. For statistical evaluation, we relied on three measures of performance: the out-of-bag (OOB) score, the area under the receiver operating curve (AUC), and the root mean square error (RMSE).

From a bagged model like a random forest, the OOB score is calculated by assessing the error of each estimator in the ensemble. Since each estimator is trained on a random subset of the training data, a prediction is made with the training samples not used to train that estimator, and the classification accuracy is calculated. For a random forest regressor such as the one we employed, the classification accuracy is assessed with the r^2 (coefficient of determination). The final OOB score is the average r^2 of all estimators (Breiman 2001), ranging from 0 to 1 where an OOB score

of 1 indicates a perfect classifier (Breiman 1996b). The OOB score is most useful as a first order estimate of model performance because it is calculated during the model training process and does not require generating a final prediction. Because of this, the OOB score could not be assessed using the geographically stratified subsampling scheme described above, but it still provides useful information on model fit.

The AUC score is a more formal evaluation method for assessing performance and requires testing predicted values against reserved testing samples. AUC is a measure of a model's overall discriminatory ability. It is derived from comparing the rate of false presences to true presences, as the cutoff changes for what is considered a positive or negative classification. AUC ranges from 0 to 1, where 1 indicates a perfect classifier and a value less than 0.5 indicates a classifier that is worse than random (Hosmer, Lemeshow, and Sturdivant 2013).

While AUC only addresses the frequency of error, RMSE is useful for assessing the magnitude. Ji and Gallo (2006) and Riemann and others (2010) contend that RMSE is less useful than root mean square percentage error because RMSE is dependent on the units of the data and is consequently not bounded or scaled. Our prediction maps ranged from 0 to 1, representing the probability of occurrence. In our case, then, RMSE also ranges from 0 to 1, and can therefore be easily understood as percent of possible error. Because RMSE decreases and the AUC and the OOB score increase with model accuracy, we subtracted RMSE from 1 to maintain a consistent scale for all metrics.

In addition to assessing agreement across the entire study area, we stratified both AUC and RMSE by land cover type. To do so, we calculated both metrics for each land cover class in the land cover predictor data, although we only report the performance for land cover types relevant to species we selected. Stratifying assessments in this way offered information about where models were making errors and how those errors relate to specific attributes of each species' distribution and temporal dynamics of different land cover types.

Lastly, we used information on variable importance to assess model structure and dynamics. In the process of constructing decision-tree-based models, the relative importance of each predictor is easy to compute. Relative importance for each predictor is determined per estimator by calculating the increase in the model's discriminatory ability (i.e., the information gain) after including that predictor in the model. For each predictor variable, importance scores are averaged for all estimators to determine importance for the entire ensemble. Importance ranges from 0 to 1, where a theoretical score of 1 indicates that predictor was the only variable that allowed the model to split all training samples. Although there are some limitations to calculating variable importance with this method (Strobl et al. 2007), importance remains an informative byproduct of decision-tree-based models.

Results

In general, model performance for all species was clearly influenced by temporal resolutions, but the best performing resolutions did not agree between species. Western Meadowlark, Hermit Thrush, and Barn Swallow models seemed least affected by changes in temporal resolution, while Black-headed Grosbeak models were most sensitive. RMSE was generally less responsive

than AUC, but both showed similar patterns of performance across all resolutions for each species except Hermit Thrushes.

To compare models of different temporal resolutions, coarser resolution models used land cover time series with larger time steps than finer resolution models. Yearly observations that did not match a year of predictor data were related to the closest year. For three species, Black-headed Grosbeaks, Barn Swallows, and White-crowned Sparrows, coarser resolution models unexpectedly performed better than finer resolution models, with 10-year models outperforming all others (Figure 3.3). For these species, the 2-year models were consistently the lowest-performing SDMs, and the 5-year models were generally comparable to the 1-year models. Differences in accuracy were most apparent in the AUC scores with an average increase between 1- and 10-year models of 24%. While performance for 10-year models markedly increased over finer resolution models for all three species, the Black-headed Grosbeak 10-year model showed dramatic improvement.

Western Meadowlark and Hermit Thrush SDMs, however, did not follow the same pattern. For Meadowlarks, the 2-year model was most accurate, but the 1-year model was comparable, with only a slightly lower AUC score. Meadowlark RMSE remained nearly constant for all resolutions with a decrease of only 0.5% between 1- and 10-year models. For Hermit Thrush models, the 5-year resolution had the highest AUC score, and the 1-year model had the lowest. According to RMSE, however, the 2-year model performed the best, and the 10-year model was the worst.

Unlike AUC and RMSE, the OOB score decreased slightly with each coarser resolution for all species. Although the decrease was subtle, it was monotonic for nearly all species. Since the OOB score is derived from training samples, it is essentially a measure of how well a model is fit to its training data, but not necessarily a reflection of a model's ability to generalize across the modeling region. In other words, the slight decrease in OOB score shows that the models with coarser temporal resolutions were less fit for making predictions at locations the models had already seen.

Model performance per land cover class (Figure 3.4) did not follow a consistent pattern between species, but it did tend to follow the same patterns as overall performance for each species. There were a few exceptions to this, however, where accuracy in certain land cover types contradicted the overall accuracy trends. For instance, both AUC and RMSE for Black-headed Grosbeak models dramatically improved in all land cover types with the 10-year model except for deciduous forest. For this land cover, AUC showed an overall decline with decreasing temporal resolution, and RMSE was highest for the 2-year model, but was otherwise constant. For Hermit Thrush models, the majority of testing samples occurred in urban land cover, which showed a slight decline in AUC with resolution below 2-years. That these urban samples did not seem to exert more influence on overall AUC scores is surprising. Similarly, White-crowned Sparrow AUC scores were highest in urban land cover with the 2-year model, whereas overall AUC was lowest for this resolution. Here, too, these urban samples did not have a strong influence on overall AUC.

Variable importance also revealed some interesting relationships between model behavior and species dynamics. In general, constant predictors (i.e., those that we did not adjust temporal resolution for) provided more information than yearly predictors (i.e., land cover), but the year that the observations were made was most frequently the most important predictor. A pattern shared by all species was an increased reliance on constant predictors with decreasing temporal resolution, and consequently, less reliance on yearly land cover predictors (Figure 3.6). In other words, yearly predictors provided less useful information to be able to distinguish between positive and negative observations. This pattern was consistent across all species, regardless of the optimal temporal resolution for that species. In comparing relative importance of land cover predictors, those processed with a 13.5 ha neighborhood consistently provided greater information gains than smaller neighborhoods and predictors at the native 30 m pixel size.

Discussion

The aim of this study was to examine the effects of temporal resolution on SDMs. We did so by training multiple models with yearly bird observations from an 11-year time period, but using land cover predictor data from years that were increasingly sparse. This meant that models of coarser resolutions were potentially trained with observations that did not match the contemporaneous land cover. We tested different temporal resolutions for five species to capture how differences in species-environment relationships may attenuate or amplify the effects of resolution. We expected some variation in SDM response to temporal resolution for different species, but also that finer temporal resolutions would generally perform better for all species. We also anticipated that SDM response would be vary by habitat preference (i.e., forest or anthropogenic landscapes). Our results, however, did not generally conform to these

expectations. Instead, we found some SDMs improved at the coarsest resolution of 10 years, while others performed better at moderate resolutions of 2 or 5 years. Among different species, we also found few discernable patterns related to habitat preference. These results raise several intriguing questions about the temporal dynamics of SDMs and their relationship to land cover, which we address below.

Potential sources of error

As with any predictive modeling task, it is important to consider the sources of error that may have contributed to the responses of SDMs to temporal resolution that we report here. Although every response dataset introduces some error to the modeling process, error present in citizen science data may be markedly larger and more frequent than data collected by experts (Yu, Wong, and Kelling 2014). While many studies have demonstrated that citizen science data are of high enough quality to use for developing SDMs (e.g., Abolafya et al. 2013; Fink et al. 2010; Johnston et al. 2015; Nagy et al. 2012), uncertainty in identification and positional accuracy poses potential barriers for accurate estimations of species occurrence.

A known challenge with using eBird data, in particular, for any kind of spatial analysis is the inevitable spatial bias in the data (Hochachka and Fink 2012). Rather than adhering to a more rigorous protocol, crowd-sourced data collection efforts are almost inevitably going to be opportunistic. This presented several challenges in our study, including adjusting our sampling scheme and choosing performance metrics that adequately reflected the models' overall capabilities to generalize for locations not seen in the training data. Although our test samples did not share any common locations with the training data, and we attempted to equitably draw

test samples from across the modeling region, the inherent spatial bias in eBird data still might have prevented our evaluations from adequately assessing predictions for locations where eBird participants are less likely to make observations.

In addition to accurate response data, adequately estimating species-environment relationships in an SDM requires predictor data that sufficiently represent environmental conditions. Our SDMs relied heavily on land cover maps for that purpose, but these maps, too, introduced error to the modeling process. This was most apparent in misclassification of land cover in certain areas. For instance, many observations of Western Meadowlarks coincided with pixels classified as deciduous forest. Meadowlarks are notoriously open-space and grassland birds, frequenting agricultural areas. Some of the Meadowlark observations from deciduous forest pixels were actually made in agricultural areas that were misclassified in the land cover data. Such commission errors could have obscured the relationships between bird observations and environmental conditions. Nevertheless, these data are the only available land cover maps that cover the entire spatial and temporal extents of our modeling region, and overall, they represent actual landscape patterns sufficiently well.

SDM response to temporal resolution

For SDMs that decreased in accuracy with coarser temporal resolutions (i.e., Western Meadowlark and Hermit Thrush models), intentionally coarsening the resolution may have had the effect of introducing corrupt samples to the training set, as many training samples offered erroneous representations of species-environment relationships. In spatial modeling, a sample is considered corrupt if it contains error such that it misrepresents relationships between predictors

and a response variable. Decision trees are highly susceptible to perturbations in the training data, which is the main reason why bagging decision trees improves predictions (Breiman 1996b). In theory, aggregating over numerous decision trees trained on bootstrap samples should improve predictions on a corrupt training set (Deitterich 2000), but there are other non-parametric models that are more stable. Support vector machines, for instance, are known for being relatively immune to moderate training sample corruption (Huang, Davis, and Townshend 2002). Such models may not be affected by the moderate changes in training data that coarse predictor temporal resolutions introduce. When modeling distributions for Western Meadowlarks, Hermit Thrushes, and other species for which more contemporaneous predictors may yield better predictions, more stable models like support vector machines may be better suited if the available predictor data do not match the years of observations.

Corrupt samples can only be considered corrupt, however, if they erroneously represent relationships between predictors and a response variable. If altering predictor values improves an SDM, it is reasonable to assume that the altered predictor values are more representative of species-environment relationships for that species. For species with higher accuracy from coarse-resolution models, adjusting the temporal resolution—instead of corrupting the samples—may have more accurately represented temporal scales important to these species with respect to land cover. Indeed, Knick and Rotenberry (2000) and Ernoult and others (2005) both demonstrated that for some bird species, past landscape patterns might be more related to species occurrence than present patterns, suggesting a time lag between landscape change and species response. Attributing the performance of coarse resolution to a time lag response in some species for this study, however, ignores the occasions when observations were related to land cover more recent

than the observations. Still, time lags may have at least partially contributed to model performance. One way to test this would be to explicitly increase the time lag (rather than temporal resolution as we have expressed it here) between observations and predictors. Although such a test would be outside the scope of this study, understanding time lag responses of varying bird species remains an important area for future research (Vierling et al. 2014).

Another potential explanation for why coarser-resolution models performed better for certain species may have had less to do with ecological relationships and more to do with the training data and the model algorithm we used. As described above, random forest models rely on two sources of randomization to create a diverse ensemble of decision trees: bagging and the random subspace method. Mechanically, this results in a model that can ignore uninformative training samples and their relationships to predictor variables and exploit more informative ones.

Additionally, the temporal distribution of the ERD is unequal, with an increasing number of observations each year (Figure 3.7). Since there are far more samples in later years, it is conceivable that if a model adequately represents samples from later years, the model will attain high accuracy scores. In concert with a random forest's tendency to gravitate toward patterns that provide the greatest discriminatory ability, it is possible that misrepresenting observations from earlier years might have boosted the importance of patterns from later years. With coarser resolution for all species, models relied especially heavily on the year that the observation was made. This is likely because the year allowed the models to separate later observations from earlier ones. To test this hypothesis, we trained a Black-headed Grosbeak model using only observation from 2010–2012 and using yearly predictors. This, however, yielded a model with an AUC of only 0.61, much lower than the AUC of 0.98 achieved by the 10-year model. The

underlying mechanisms leading to more accurate models with coarser resolutions remain unknown and represent an area in need of further study.

Regardless of the reason why coarser temporal resolution models performed better for certain species, the occurrence maps themselves support the results from statistical evaluation. It is reasonable to assume that within a single map, occurrence probabilities from pixels near to one another should be more similar than pixels further away. Maps where nearby pixels are dissimilar to one another are likely indicative of a noisy model with poor discriminatory ability, as such a model may not have been able to discern broad-scale patterns that define distributions across large extents. Summarizing the agreement among pixels in the same area with the standard deviation is one way to assess this. Figure 3.8 shows the difference in standard deviations within 35,000 ha hexagons between the 10-year Black-headed Grosbeak occurrence map and the 1-year map. Negative values occur where the 1-year standard deviation was higher than the 10-year deviation. With an overwhelming number of negative values, it is clear that the 10-year model's discriminatory capabilities were higher. White-crowned sparrow and Barn Swallow occurrence maps show a similar reduction of noise in the 10-year maps.

SDM response across habitat preferences

Our intent in modeling distributions for multiple species was to see if habitat preference was related to an SDM's response to temporal resolution. We found little evidence of such relationships according to habitat preferences, but model performance for different species did show subtle signs of such an effect. This was only evident in the magnitude of changes in accuracy with different temporal resolutions for each species. AUC changed much less for both

species commonly found in agricultural and urban landscapes (i.e., Barn Swallows and Western Meadowlarks) than it did for forest-adapted species. The average difference in AUC for these two species between the worst and best models was 0.09, whereas the average change for Black-headed Grosbeaks, Hermit Thrushes, and White-crowned Sparrows was 0.25.

It is important to note, as well, that Hermit Thrushes model performance was much less affected by temporal resolution, despite its associations with forest habitats. This is likely a product of the relatively few positive observations of this species and the effect that had on statistical evaluation. The low number of positive observations of this species resulted in an occurrence map with far more pixels with low occurrence probability values. Hermit Thrush AUC only changed by 0.08 because most testing samples with a value of 0 agreed with the maps' low pixel values. This meant that it was more likely for this species that models from each temporal resolution would statistically agree with training samples than likelihood of agreement for other forest birds. Hence, there was far less variation in models of different temporal resolutions.

In addition, White-crowned Sparrows are not necessarily a forest bird, and in the non-breeding season they are common in agricultural and urban landscapes. However, in the breeding season, while most individuals migrate north to Canada and Alaska, some isolated populations remain in coniferous forests in California, Oregon, and Washington, frequenting disturbed area such as clear-cuts. Sparrows that remain in the conterminous United States in the breeding season are, therefore, likely to be affected by changes in these forests. Although this species is not truly a forest bird, for the time of year for which we considered bird distributions, it is appropriate to consider its relationship to this land cover type.

Still, the comparatively small response of Barn Swallow and Western Meadowlark models to temporal resolution might have resulted from the relative inter-annual stability of the urban and agricultural landscapes in which both species are commonly found. Although intra-annual temporal patterns for agricultural lands can be quite dynamic, agriculture and urban land covers tend to be more consistent year to year than other land covers like forests (Homer et al. 2015). Since both Swallows and Meadowlarks are common in these land covers, it is conceivable that changes in temporal resolution would have less of an effect on these species. For agricultural landscapes, it is difficult to confirm this from a comparison of model performance by land cover type (Figure 3.4) because these landscapes are not explicitly represented as a single land cover class in the data we used; agricultural land is a designation of land use rather than explicitly land cover. Instead, agricultural lands are represented by a mixture of mostly bare ground and shrub, with some misclassified deciduous forest. As such, there is little discernable pattern in these land covers as they relate to agriculture. In urban landscapes, Barn Swallow AUC remained fairly consistent with a change of only 0.05 between the best and worst models, and Western Meadowlark models only differed by 0.08 in AUC. White-crowned sparrows and Hermit Thrushes also demonstrated similar consistency in urban areas across temporal resolutions, so this pattern is not necessarily unique to the two urban and agriculturally adapted species (i.e., Barn Swallows and Western Meadowlarks).

Considering that differences in the magnitude of responses could not definitively be separated by habitat preferences, it is worth noting that responses to disturbances and other changes in environmental conditions are notoriously inconsistent among different bird species (Knick and

Rotenberry 2000; Simon, Schwab, and Otto 2002; Stephens and Alexander 2011). Considering that we have only developed models for five species, generalizations about habitat preference and its relationship with responses to temporal resolution are likely difficult to make, if not inaccurate. The most compelling reason to make such generalizations would be to aid the selection of an appropriate temporal resolution for another species that may exhibit similar habitat preferences. Rather than making that choice from some generalized relationship, however, consideration of a combination of bird behavior (e.g., sensitivity to disturbance or site fidelity) and the qualities of the landscape across a modeling region (e.g., land cover diversity or rates and magnitudes of disturbance) are likely to produce a better-specified model.

Model performance across different evaluation metrics

The use of multiple evaluation metrics is critical when assessing predictive models, since each metric offers different information about errors present in a model (Foody 2002; Riemann et al. 2010; Strahler et al. 2006). In this study we used AUC and RMSE to provide information on the frequency and magnitude of errors. The OOB score was also useful for indicating how well the models were fit to their training data.

Perhaps one of the most compelling reasons to use multiple assessment methods is that some evaluation metrics might present overly optimistic results, depending on the nature of the training data or the prediction. Models trained on data that are spatially biased, such as eBird, may be particularly susceptible to this. In our study, the OOB score presented the same trend of model performance for all species decreasing with coarser resolution, regardless of the performance according to AUC and RMSE. Breiman (1996a) introduced the OOB score as an

efficient and unbiased estimation of model accuracy for bagged models. This assertion, however, assumes an unbiased training set, with the probability that two random samples from the training set share no more of a relationship than a single sample from the training set and a single sample from the test set. As mentioned above, the OOB score is a measure of how well a model is fit to its training data, information that could be useful for calibrating model parameters regardless of any spatial biases present. For tasks such as this, the OOB score may still have some utility for developing models with spatially biased datasets, but model assessment using the OOB score should be selective.

The diverse information that multiple metrics provide can also reveal systematic errors that would otherwise go unnoticed with only a single metric. For instance, Hermit Thrush AUC and RMSE showed opposing responses to temporal resolution, with AUC suggesting increased performance with coarser resolutions and RMSE indicating increased performance with finer resolutions. Although all models predicted absences better than presences, this difference was greatest for Hermit Thrush models. For this species, this particularly uneven performance resulted in a discrepancy between AUC and RMSE, with AUC showing the 5-year model performing the best and RMSE showing the 2-year model performing the best. This discrepancy also points to the different information gleaned from each metric. Hermit Thrush occurrence probabilities were lower on average than for any other species, so the potential magnitude of error was relatively high for positive observations. As such, RMSE was sensitive to false absences, and it indicated that coarse-resolution models were less capable of predicting presences than absences. Since AUC is a measure of the frequency of error and there were far more absences reported, AUC largely ignored the increased rate of false absences. Whether

coarse resolution or fine resolution models ultimately performed better for this species depends on whether predicting presences or absences is more important for a particular application of an SDM. Either way, the differences in AUC and RMSE results underscore the importance of using multiple assessment metrics when statistically comparing model products.

Conclusion

A central concern in ecology is determining the scales important to a particular phenomenon. Spatially, these can be understood in part as the distance beyond which the relationships between a phenomenon and the environmental conditions that affect it have changed. Temporally, the question is the same: how much time can pass before the state of a phenomenon is no longer related to some set of original environmental conditions?

We attempted to address this question by training avian SDMs with land cover predictors that were increasingly temporally sparse, essentially comparing the performance of SDMs of different temporal resolutions. This meant that for many observations of birds for coarse-resolution models, land cover at those locations was not contemporaneously represented. Since land cover is strongly related to bird distributions (Manel et al. 2000; Mitchel et al. 2001; Rullman and Marzluff 2014), we anticipated that coarser temporal resolution would produce a less accurate model. For two species, this held true. For the three other species considered here, the coarsest resolution produced the most accurate models.

These results suggest that temporal interactions between bird occurrence and land cover are more complex than we originally anticipated. Previous research demonstrates that the responses to

disturbances vary among different species (Simon, Schwab, and Otto 2002; Stephens and Alexander 2011). Some species may also show a time lag between the timing of a disturbance and use of a site, while others are more likely to respond to current environmental conditions (Ernault et al. 2005; Knick and Rotenberry 2000). Much like spatial scales, however, there are likely multiple temporal scales important to a given species. Including multi-temporal predictors for each observation of a training set could prove to be an effective way to represent temporal patterns for certain species.

The potential influence of lag time and the degree to which it influenced our results was further complicated by the uneven temporal distribution of the training data we used. Because of this, models might have favored certain temporal patterns over others. We decided to use eBird observations as training data because of the broad spatial and temporal extent combined with fine resolution that they offer. Reducing the extent of our modeling region, however, may have been more appropriate for addressing our central question about temporal resolution. With a smaller extent, other bird occurrence datasets without this temporal bias might have been appropriate.

This research represents one of only a few studies of temporal resolution and its effect on SDMs. Our complex results demonstrate this will likely be a fruitful area for future research. Further investigation of temporal resolution and its relationship with not only other bird species, but also other taxa could reveal other unexpected species-environment relationships. Our models relied on unique yearly land cover, but such multi-temporal datasets are becoming increasingly common. This will hopefully make analyses of temporal resolution more accessible to other lines of questioning.

References

- Abolafya, Moris, Ortaç Onmuş, Çağan H. Şekercioğlu, and Raşit Bilgin. 2013. “Using Citizen Science Data to Model the Distributions of Common Songbirds of Turkey Under Different Global Climatic Change Scenarios.” Edited by Vanesa Magar. *PLoS ONE* 8 (7): e68037. doi:10.1371/journal.pone.0068037.
- Addicott, John F., John M. Aho, Michael F. Antolin, Dianna K. Padilla, John S. Richardson, and Daniel A. Soluk. 1987. “Ecological Neighborhoods: Scaling Environmental Patterns.” *Oikos* 49 (3): 340–46. doi:10.2307/3565770.
- Adler, Peter B., and William K. Lauenroth. 2003. “The Power of Time: Spatiotemporal Scaling of Species Diversity.” *Ecology Letters* 6 (8): 749–56. doi:10.1046/j.1461-0248.2003.00497.x.
- Araújo, Miguel B., and Mark New. 2007. “Ensemble Forecasting of Species Distributions.” *Trends in Ecology & Evolution* 22 (1): 42–47. doi:10.1016/j.tree.2006.09.010.
- Ashbaugh, James G, and John A Alwin. 1994. *The Pacific Northwest: Geographical Perspectives*. Dubuque, Ia.: Kendall/Hunt Pub. Co.
- Austin, Mike P., and Kimberly P. Van Niel. 2011. “Improving Species Distribution Models for Climate Change Studies: Variable Selection and Scale.” *Journal of Biogeography* 38 (1): 1–8. doi:10.1111/j.1365-2699.2010.02416.x.
- Betts, Matthew G., A.W. Diamond, G.J. Forbes, M.-A. Villard, and J.S. Gunn. 2006. “The Importance of Spatial Autocorrelation, Extent and Resolution in Predicting Forest Bird Occurrence.” *Ecological Modelling* 191 (2): 197–224. doi:10.1016/j.ecolmodel.2005.04.027.
- Boakes, Elizabeth H., Philip J. K. McGowan, Richard A. Fuller, Ding Chang-qing, Natalie E. Clark, Kim O’Connor, and Georgina M. Mace. 2010. “Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data.” *PLoS Biol* 8 (6): e1000385. doi:10.1371/journal.pbio.1000385.
- Breiman, Leo. 1984. *Classification and Regression Trees*. The Wadsworth Statistics/Probability Series. Belmont, Calif: Wadsworth International Group.
- . 1996a. “Bagging Predictors.” *Machine Learning* 24 (2): 123–140.
- . 1996b. “Heuristics of Instability and Stabilization in Model Selection.” *The Annals of Statistics* 24 (6): 2350–83.

- . 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.
- Bulluck, Lesley, Erica Fleishman, Chris Betrus, and Rob Blair. 2006. “Spatial and Temporal Variations in Species Occurrence Rate Affect the Accuracy of Occurrence Models.” *Global Ecology and Biogeography* 15 (1): 27–38. doi:10.1111/j.1466-822X.2006.00170.x.
- Cohen, Warren B., and Samuel N. Goward. 2004. “Landsat’s Role in Ecological Applications of Remote Sensing.” *BioScience* 54 (6): 535–45. doi:10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2.
- Cohen, Warren B., Zhiqiang Yang, and Robert Kennedy. 2010. “Detecting Trends in Forest Disturbance and Recovery Using Yearly Landsat Time Series: 2. TimeSync — Tools for Calibration and Validation.” *Remote Sensing of Environment* 114 (12): 2911–24. doi:10.1016/j.rse.2010.07.010.
- Dickinson, Janis L., Benjamin Zuckerberg, and David N. Bontar. 2010. “Citizen Science as an Ecological Research Tool: Challenges and Benefits.” *Annual Review of Ecology, Evolution, and Systematics* 41 (1): 149–72. doi:10.1146/annurev-ecolsys-102209-144636.
- Diefenbach, Duane R., Matthew R. Marshall, Jennifer A. Mattice, Daniel W. Brauning, and D. H. Johnson. 2007. “Incorporating Availability for Detection in Estimates of Bird Abundance.” *The Auk* 124 (1): 96–106.
- Dietterich, Thomas G. 2000. “An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization.” *Machine Learning* 40 (2): 139–157.
- Elith, Jane, and John R. Leathwick. 2009. “Species Distribution Models: Ecological Explanation and Prediction Across Space and Time.” *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 677–97. doi:10.1146/annurev.ecolsys.110308.120159.
- Ernault, Aude, Yves Tremauville, Dominique Cellier, Pierre Margerie, Estelle Langlois, and Didier Alard. 2006. “Potential Landscape Drivers of Biodiversity Components in a Flood Plain: Past or Present Patterns?” *Biological Conservation* 127 (1): 1–17. doi:10.1016/j.biocon.2005.07.008.
- Ferraz, Katia Maria P. M. de Barros, Silvio Frosini de Barros Ferraz, Rogerio Cunha de Paula, Beatriz Beisiegel, and Christine Breitenmoser. 2012. “Species Distribution Modeling for Conservation Purposes.” *Natureza & Conservação* 10 (2): 214–20. doi:10.4322/natcon.2012.032.
- Fink, Daniel, Wesley M. Hochachka, Benjamin Zuckerberg, David W. Winkler, Ben Shaby, M. Arthur Munson, Giles Hooker, Mirek Riedewald, Daniel Sheldon, and Steve Kelling. 2010. “Spatiotemporal Exploratory Models for Broad-Scale Survey Data.” *Ecological Applications* 20 (8): 2131–2147.

- Foody, Giles M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80 (1): 185–201.
- Franklin, Jerry F., and C. T. Dyrness. 1988. *Natural Vegetation of Oregon and Washington*. Portland, Or., Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Dept. of Agriculture. <http://ir.library.oregonstate.edu/xmlui/handle/1957/34706>.
- Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. "The National Elevation Dataset." *Photogrammetric Engineering and Remote Sensing* 68 (1): 5–+.
- Gottschalk, Thomas K., Birgit Aue, Stefan Hotes, and Klemens Ekschmitt. 2011. "Influence of Grain Size on Species–habitat Models." *Ecological Modelling* 222 (18): 3403–12. doi:10.1016/j.ecolmodel.2011.07.008.
- Gray, Josh, and Conghe Song. 2013. "Consistent Classification of Image Time Series with Automatic Adaptive Signature Generalization." *Remote Sensing of Environment* 134 (July): 333–41. doi:10.1016/j.rse.2013.03.022.
- Gschweng, Marion, Elisabeth K. V. Kalko, Peter Berthold, Wolfgang Fiedler, and Jakob Fahr. 2012. "Multi-Temporal Distribution Modelling with Satellite Tracking Data: Predicting Responses of a Long-Distance Migrant to Changing Environmental Conditions." *Journal of Applied Ecology* 49 (4): 803–13. doi:10.1111/j.1365-2664.2012.02170.x.
- Gutzwiller, Kevin J., and Wylie C. Barrow. 2001. "Bird–Landscape Relations in the Chihuahuan Desert: Coping with Uncertainties About Predictive Models." *Ecological Applications* 11 (5): 1517–32. doi:10.1890/1051-0761(2001)011[1517:BLRITC]2.0.CO;2.
- Hejl S. J., E. C. Beedy. 1986. "Weather induced variation in the abundance of birds." In Verner, Jared, Michael L. Morrison, C. John Ralph, Pacific Southwest Forest and Range Experiment Station (Berkeley, Calif.), and Wildlife Society, eds. *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates: Based on an International Symposium Held at Stanford Sierra Camp, Fallen Leaf Lake, California, 7-11 October 1984*. Madison, Wis: University of Wisconsin Press, 1986.
- Ho, Tin Kam. 1998. "The Random Subspace Method for Constructing Decision Forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (8): 832–844.
- Hochachka, Wesley, and Daniel Fink. 2012. "Broad-Scale Citizen Science Data from Checklists: Prospects and Challenges for Macroecology." *Frontiers of Biogeography* 4 (4). <http://escholarship.org/uc/item/01t5c00w>.
- Homer, Collin, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Xian, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. 2015. "Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information." *Photogrammetric Engineering & Remote Sensing* 81 (5): 345–54. doi:10.1016/S0099-1112(15)30100-2.

- Hosmer, David W., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression*. Third edition. Wiley Series in Probability and Statistics. Hoboken, New Jersey: Wiley.
- Huang, C., L. S. Davis, and J. R. G. Townshend. 2002. "An Assessment of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 23 (4): 725–49. doi:10.1080/01431160110040323.
- Huang, Chengquan, Samuel N. Goward, Jeffrey G. Masek, Nancy Thomas, Zhiliang Zhu, and James E. Vogelmann. 2010. "An Automated Approach for Reconstructing Recent Forest Disturbance History Using Dense Landsat Time Series Stacks." *Remote Sensing of Environment* 114 (1): 183–98. doi:10.1016/j.rse.2009.08.017.
- Ives, Anthony R., and Eric D. Klopfer. 1997. "Spatial Variation in Abundance Created by Stochastic Temporal Variation." *Ecology* 78 (6): 1907–1913.
- Ji, Lei, and Kevin Gallo. 2006. "An Agreement Coefficient for Image Comparison." *Photogrammetric Engineering & Remote Sensing* 72 (7): 823–833.
- Johnson, Douglas H. 2008. "In Defense of Indices: The Case of Bird Surveys." *Journal of Wildlife Management* 72 (4): 857–68. doi:10.2193/2007-294.
- Johnson, Ned K. 1995. "Seven Avifaunal Censuses Spanning One-Half Century on an Island of White Firs (*Abies Concolor*) in the Mojave Desert." *The Southwestern Naturalist* 40 (1): 76–85.
- Johnston, Alison, Daniel Fink, Mark D. Reynolds, Wesley M. Hochachka, Brian L. Sullivan, Nicholas E. Bruns, Eric Hallstein, Matt S. Merrifield, Sandi Matsumoto, and Steve Kelling. 2015. "Abundance Models Improve Spatial and Temporal Prioritization of Conservation Resources." *Ecological Applications* 25 (7): 1749–1756.
- Kennedy, Robert E., Zhiqiang Yang, Justin Braaten, Catharine Copass, Natalya Antonova, Chris Jordan, and Peder Nelson. 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 (September): 271–85. doi:10.1016/j.rse.2015.05.005.
- Kennedy, Robert E., Zhiqiang Yang, and Warren B. Cohen. 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. doi:10.1016/j.rse.2010.07.008.
- Kennedy, Robert E., Zhiqiang Yang, Warren B. Cohen, Eric Pfaff, Justin Braaten, and Peder Nelson. 2012. "Spatial and Temporal Patterns of Forest Disturbance and Regrowth within the Area of the Northwest Forest Plan." *Remote Sensing of Environment* 122 (July): 117–33. doi:10.1016/j.rse.2011.09.024.

- Kim, Do-Hyung, Joseph O. Sexton, Praveen Noojipady, Chengquan Huang, Anupam Anand, Saurabh Channan, Min Feng, and John R. Townshend. 2014. "Global, Landsat-Based Forest-Cover Change from 1990 to 2000." *Remote Sensing of Environment* 155 (December): 178–93. doi:10.1016/j.rse.2014.08.017.
- Knick, Steven T., and John T. Rotenberry. 2000. "Ghosts of Habitats Past: Contribution of Landscape Change to Current Habitats Used by Shrubland Birds." *Ecology* 81 (1): 220–227.
- Manel, S., S.t. Buckton, and S.j. Ormerod. 2000. "Testing Large-Scale Hypotheses Using Surveys: The Effects of Land Use on the Habitats, Invertebrates and Birds of Himalayan Rivers." *Journal of Applied Ecology* 37 (5): 756–70. doi:10.1046/j.1365-2664.2000.00537.x.
- Masek, Jeffrey G., Chengquan Huang, Robert Wolfe, Warren Cohen, Forrest Hall, Jonathan Kutler, and Peder Nelson. 2008. "North American Forest Disturbance Mapped from a Decadal Landsat Record." *Remote Sensing of Environment* 112 (6): 2914–26. doi:10.1016/j.rse.2008.02.010.
- McClintock, Brett T., Larissa L. Bailey, Kenneth H. Pollock, and Theodore R. Simons. 2010. "Unmodeled Observation Error Induces Bias When Inferring Patterns and Dynamics of Species Occurrence via Aural Detections." *Ecology* 91 (8): 2446–2454.
- Miller, Crane S., and Richard S. Hyslop. 1983. *California, the Geography of Diversity*. 1st ed. Palo Alto, Calif: Mayfield Pub. Co.
- Mitchell, Michael S., Richard A. Lancia, and John A. Gerwin. 2001. "Using Landscape-Level Data to Predict the Distribution of Birds on a Managed Forest: Effects of Scale." *Ecological Applications* 11 (6): 1692–1708. doi:10.1890/1051-0761(2001)011[1692:ULLDTP]2.0.CO;2.
- Munson, M. Arthur, Kevin Webb, Daniel Sheldon, Daniel Fink, Wesley M. Hochachka, Marshall Iliff, Mirek Riedewald, Daria Sorokina, Brian Sullivan, Christopher Wood, and Steve Kelling. "The eBird Reference Dataset, Version 5.0." Cornell Lab of Ornithology and National Audubon Society, Ithaca, NY, January 2013.
- Nagy, Christopher, Kyle Bardwell, Robert F. Rockwell, Rod Christie, and Mark Weckel. 2012. "Validation of a Citizen Science-Based Model of Site Occupancy for Eastern Screech Owls with Systematic Data in Suburban New York and Connecticut." <https://www.bioone.org/doi/full/10.1656/045.019.s611>.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (Oct): 2825–2830.
- Pielke, Roger A., Andy Pitman, Dev Niyogi, Rezaul Mahmood, Clive McAlpine, Faisal Hossain, Kees Klein Goldewijk, et al. 2011. "Land Use/Land Cover Changes and Climate: Modeling

- Analysis and Observational Evidence.” *Wiley Interdisciplinary Reviews: Climate Change* 2 (6): 828–50. doi:10.1002/wcc.144.
- PRISM Climate Group. 2012. “30-Year Normals.” Oregon State University. Corvallis, OR. <http://prism.oregonstate.edu>. Version 14.1-20140502-1000. Created 11 Jul 2012.
- Radchuk, Viktoriia, Karin Johst, Jürgen Groeneveld, Camille Turlure, Volker Grimm, and Nicolas Schtickzelle. 2014. “Appropriate Resolution in Time and Model Structure for Population Viability Analysis: Insights from a Butterfly Metapopulation.” *Biological Conservation* 169 (January): 345–54. doi:10.1016/j.biocon.2013.12.004.
- Riemann, Rachel, Barry Tyler Wilson, Andrew Lister, and Sarah Parks. 2010. “An Effective Assessment Protocol for Continuous Geospatial Datasets of Forest Characteristics Using USFS Forest Inventory and Analysis (FIA) Data.” *Remote Sensing of Environment* 114 (10): 2337–52. doi:10.1016/j.rse.2010.05.010.
- Rullman, Stan, and John M. Marzluff. 2014. “Raptor Presence Along an Urban–Wildland Gradient: Influences of Prey Abundance and Land Cover.” *Journal of Raptor Research* 48 (3): 257–72. doi:10.3356/JRR-13-32.1.
- Schroeder, Todd A., Warren B. Cohen, and Zhiqiang Yang. 2007. “Patterns of Forest Regrowth Following Clearcutting in Western Oregon as Determined from a Landsat Time-Series.” *Forest Ecology and Management* 243 (2–3): 259–73. doi:10.1016/j.foreco.2007.03.019.
- Scott, J. Michael, ed. 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Washington, DC: Island Press.
- Seo, C., J. H Thorne, L. Hannah, and W. Thuiller. 2009. “Scale Effects in Species Distribution Models: Implications for Conservation Planning under Climate Change.” *Biology Letters* 5 (1): 39–43. doi:10.1098/rsbl.2008.0476.
- Sheeren, D., S. Bonthoux, and G. Balent. 2014. “Modeling Bird Communities Using Unclassified Remote Sensing Imagery: Effects of the Spatial Resolution and Data Period.” *Ecological Indicators* 43 (August): 69–82. doi:10.1016/j.ecolind.2014.02.023.
- Simon, N Pp, F E Schwab, and R D Otto. 2002. “Songbird Abundance in Clear-Cut and Burned Stands: A Comparison of Natural Disturbance and Forest Management.” *Canadian Journal of Forest Research* 32 (8): 1343–50. doi:10.1139/x02-057.
- Song, Wonkyong, Eunyoung Kim, Dongkun Lee, Mounjiin Lee, and Seong-Woo Jeon. 2013. “The Sensitivity of Species Distribution Modeling to Scale Differences.” *Ecological Modelling* 248 (January): 113–18. doi:10.1016/j.ecolmodel.2012.09.012.
- Stephens, Jaime L., and John D. Alexander. 2011. “Effects of Fuel Reduction on Bird Density and Reproductive Success in Riparian Areas of Mixed-Conifer Forest in Southwest

- Oregon.” *Forest Ecology and Management* 261 (1): 43–49.
doi:10.1016/j.foreco.2010.09.022.
- Strahler, Alan H., Luigi Boschetti, Giles M. Foody, Mark A. Friedl, Matthew C. Hansen, Martin Herold, Philippe Mayaux, Jeffrey T. Morisette, Stephen V. Stehman, and Curtis E. Woodcock. 2006. “Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps.” *European Communities, Luxembourg* 51. http://cndwebzine.hcp.ma/cnd_sii/IMG/pdf/Document2222222222-17.pdf.
- Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. “Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution.” *BMC Bioinformatics* 8: 25. doi:10.1186/1471-2105-8-25.
- Townshend, John R., Jeffrey G. Masek, Chengquan Huang, Eric F. Vermote, Feng Gao, Saurabh Channan, Joseph O. Sexton, et al. 2012. “Global Characterization and Monitoring of Forest Cover Using Landsat Data: Opportunities and Challenges.” *International Journal of Digital Earth* 5 (5): 373–97. doi:10.1080/17538947.2012.713190.
- US Census Bureau. 2013. “2010 Census Urban Area Facts.” March 7.
<http://www.census.gov/geo/reference/ua/uafacts.html>.
- US Environmental Protection Agency. 2012. “U.S. Level III and IV Ecoregions (U.S. EPA).” Research Triangle Park, NC. <https://catalog.data.gov/dataset/u-s-level-iii-and-iv-ecoregions-u-s-epa>. Created 26 Apr 2012.
- Vaughan, I. P., and S. J. Ormerod. 2003. “Improving the Quality of Distribution Models for Conservation by Addressing Shortcomings in the Field Collection of Training Data.” *Conservation Biology* 17 (6): 1601–11. doi:10.1111/j.1523-1739.2003.00359.x.
- Vierling, Kerri T., Charles E. Swift, Andrew T. Hudak, Jody C. Vogeler, and Lee A. Vierling. 2014. “How Much Does the Time Lag between Wildlife Field-Data Collection and LiDAR-Data Acquisition Matter for Studies of Animal Distributions? A Case Study Using Bird Communities.” *Remote Sensing Letters* 5 (2): 185–93. doi:10.1080/2150704X.2014.891773.
- Wang, Jie, Yuanyuan Zhao, Congcong Li, Le Yu, Desheng Liu, and Peng Gong. 2015. “Mapping Global Land Cover in 2001 and 2010 with Spatial-Temporal Consistency at 250m Resolution.” *ISPRS Journal of Photogrammetry and Remote Sensing* 103 (May): 38–47. doi:10.1016/j.isprsjprs.2014.03.007.
- Xin, Qinchuan, Pontus Olofsson, Zhe Zhu, Bin Tan, and Curtis E. Woodcock. 2013. “Toward near Real-Time Monitoring of Forest Disturbance by Fusion of MODIS and Landsat Data.” *Remote Sensing of Environment* 135 (August): 234–47. doi:10.1016/j.rse.2013.04.002.
- Yu, Jun, Weng-Keen Wong, and Steve Kelling. 2014. “Clustering Species Accumulation Curves to Identify Skill Levels of Citizen Scientists Participating in the eBird Project.” <http://web.engr.oregonstate.edu/~wong/papers/pdf/IAAI2014.SAC.pdf>.

Zhu, Chunhao, and Yingkui Li. 2013. "High Temporal Resolution Land Use/Land Cover Change from 1984 to 2010 of the Little River Watershed, Tennessee, Investigated Using Landsat and Google Earth Images." *Southeastern Geographer* 53 (3): 250–266.

Figures

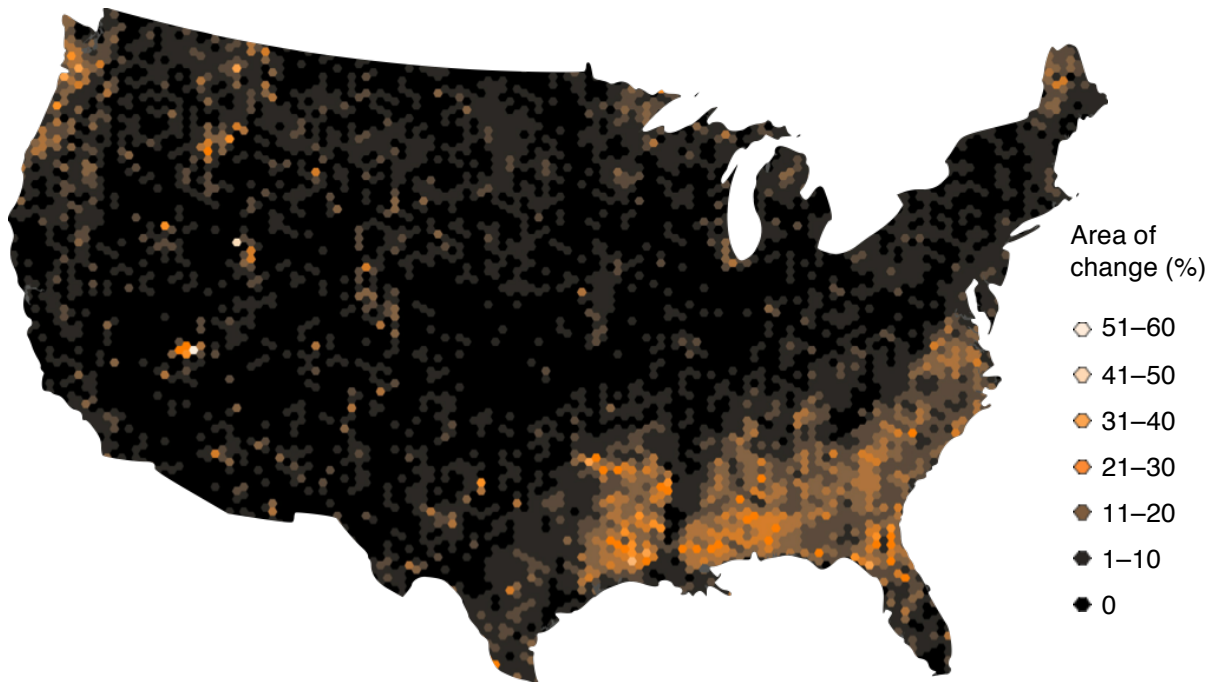


Figure 3.1. Land cover change by area between 2001 and 2011 from the National Land Cover Database (Homer et al. 2015). For display, we calculated the percent of pixels that exhibited a change of land cover within 35,000 ha hexagons.

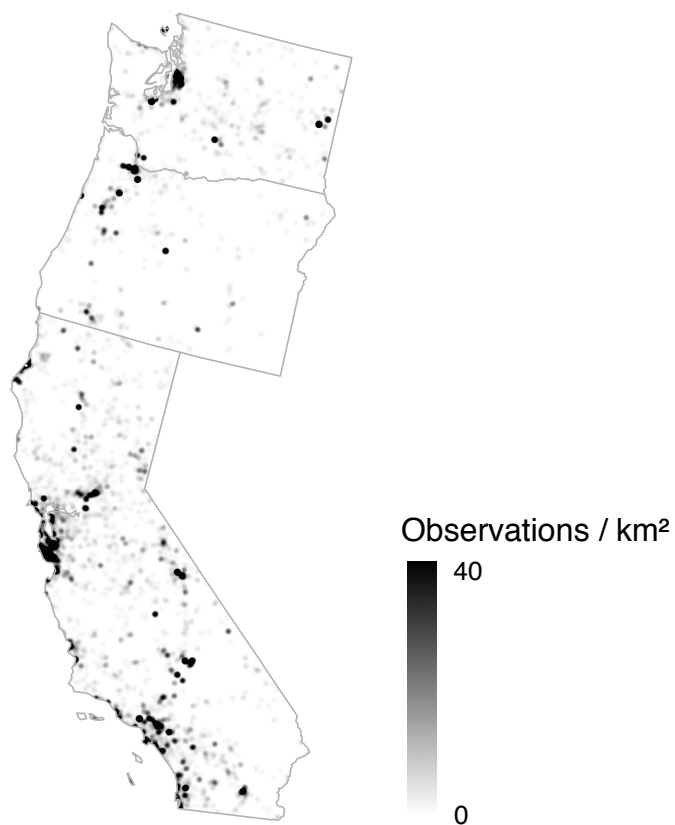


Figure 3.2. Density of ERD observations within a 1 km radius. Observations are mostly clustered around roads, trails, and population centers.

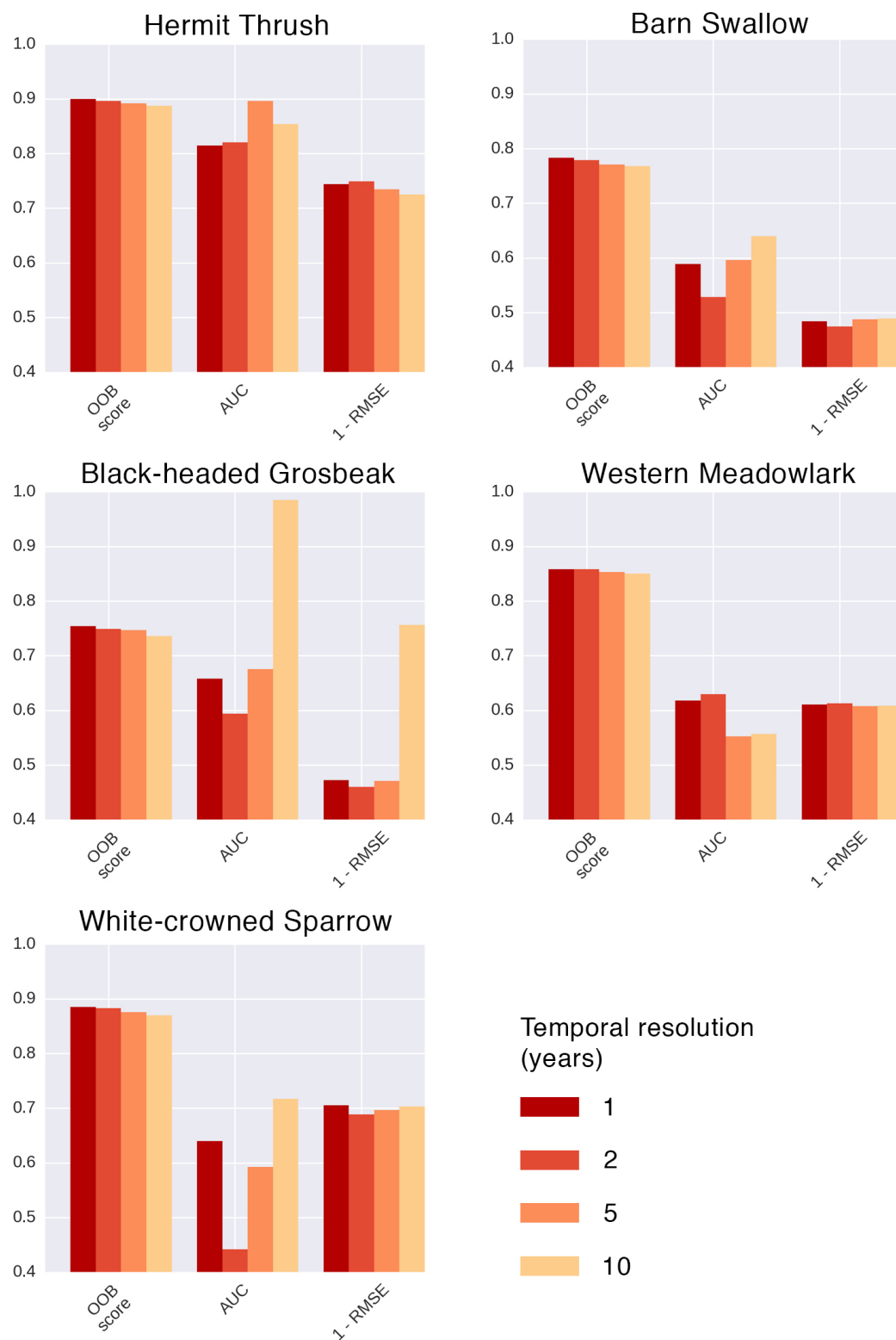


Figure 3.3. Model performance according to the OOB score, AUC, and RMSE for four different temporal resolutions.

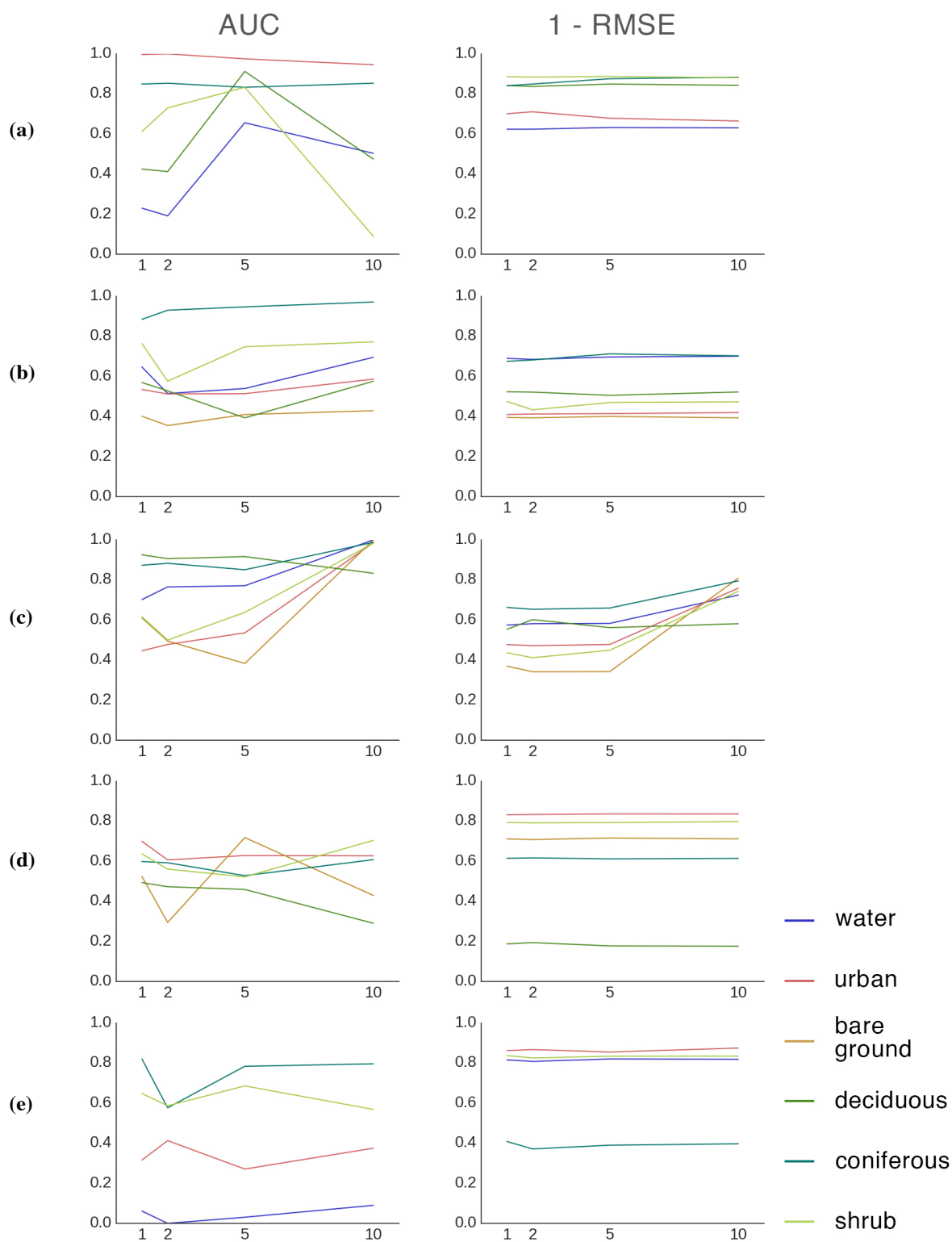


Figure 3.4. RMSE and AUC by land cover type for four different temporal resolutions for a) Hermit Thrush, b) Barn Swallow, c) Black-headed Grosbeak, d) Western Meadowlark, and e) White-crowned Sparrow.

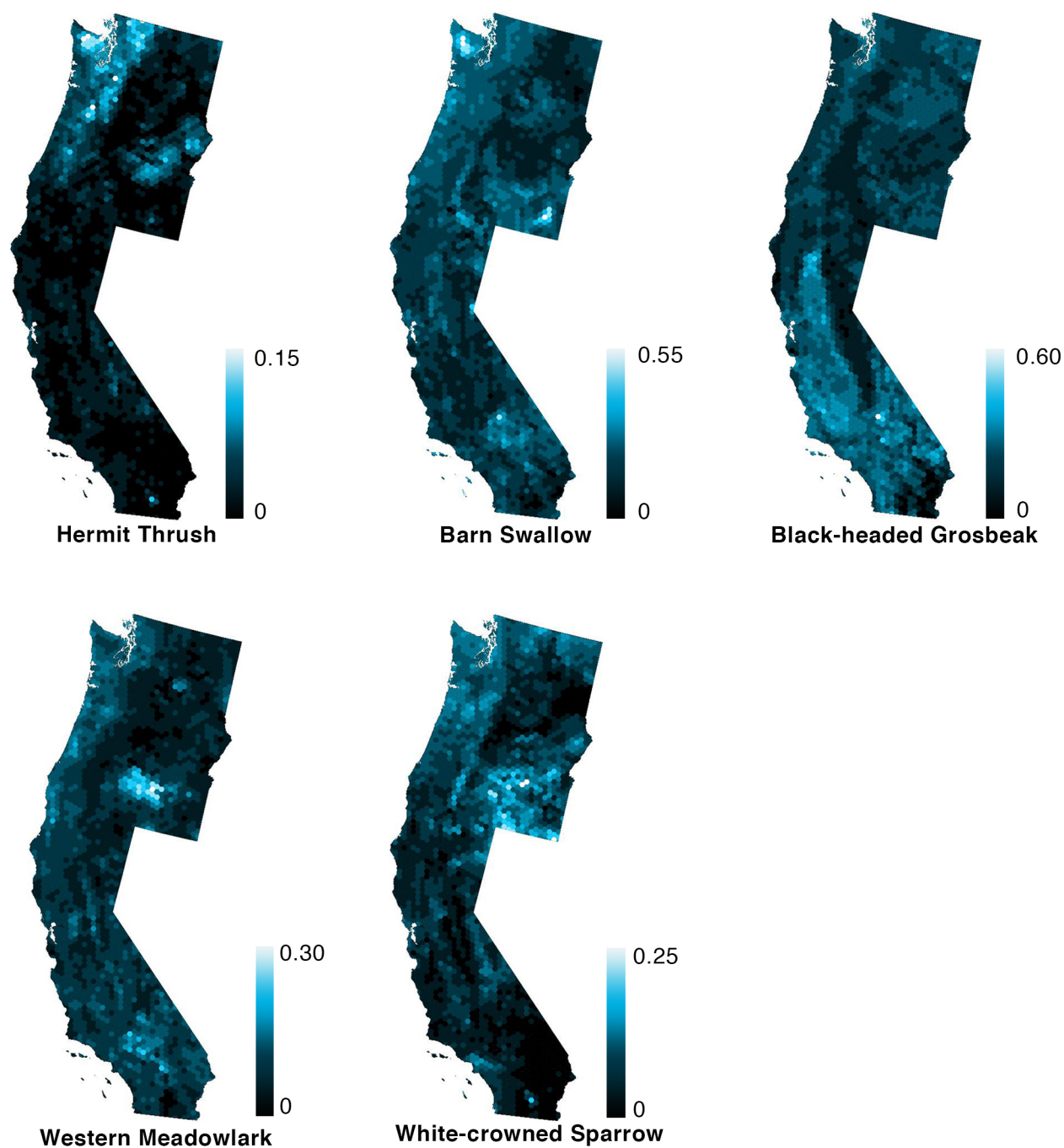


Figure 3.5. Occurrence probabilities for all five species averaged within 35,000 ha hexagons. Occurrence maps were developed with a 30 m raster cell size, but we aggregated here for visual display. Range of occurrence probabilities differed for each species, so values next to the color ramp for each map indicate the range of averaged values.

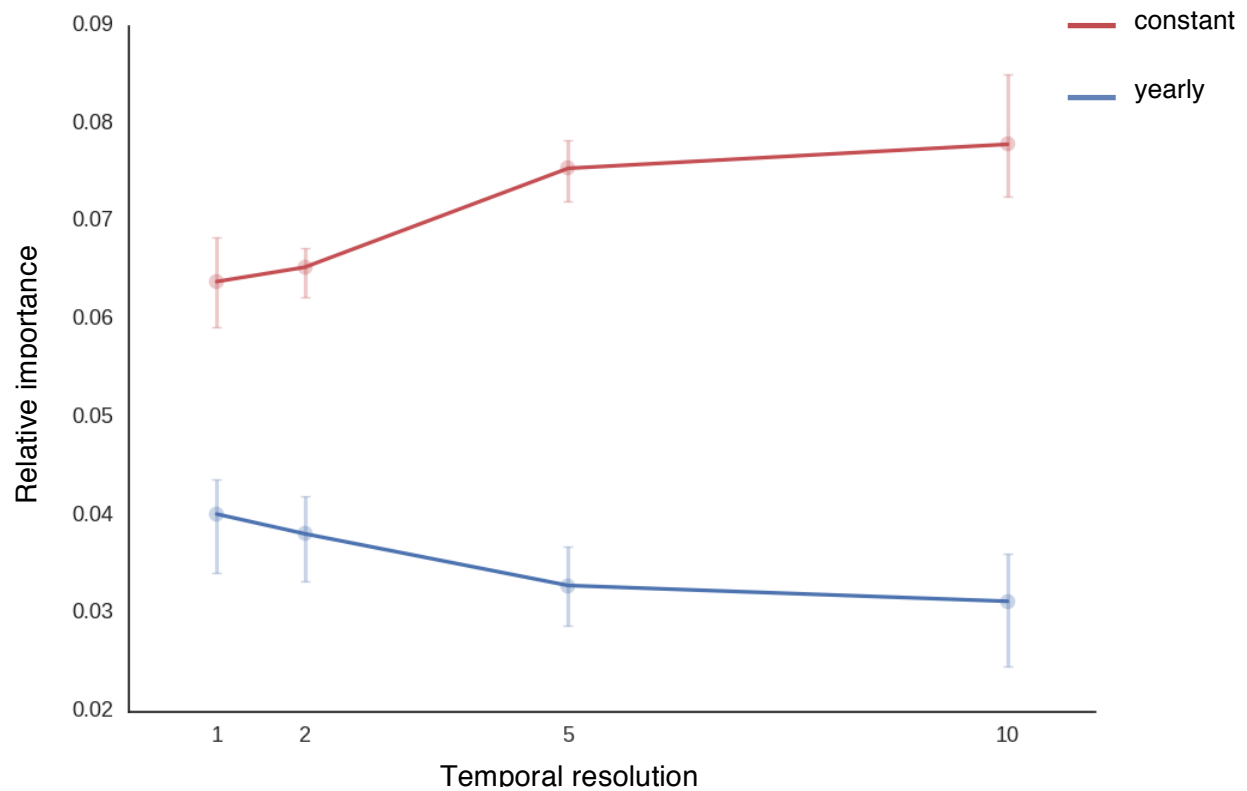


Figure 3.6. Mean relative importance for all species of constant predictors (red) and yearly land cover predictors (blue) for each temporal resolution. Vertical ticks show the range of importance for all species.

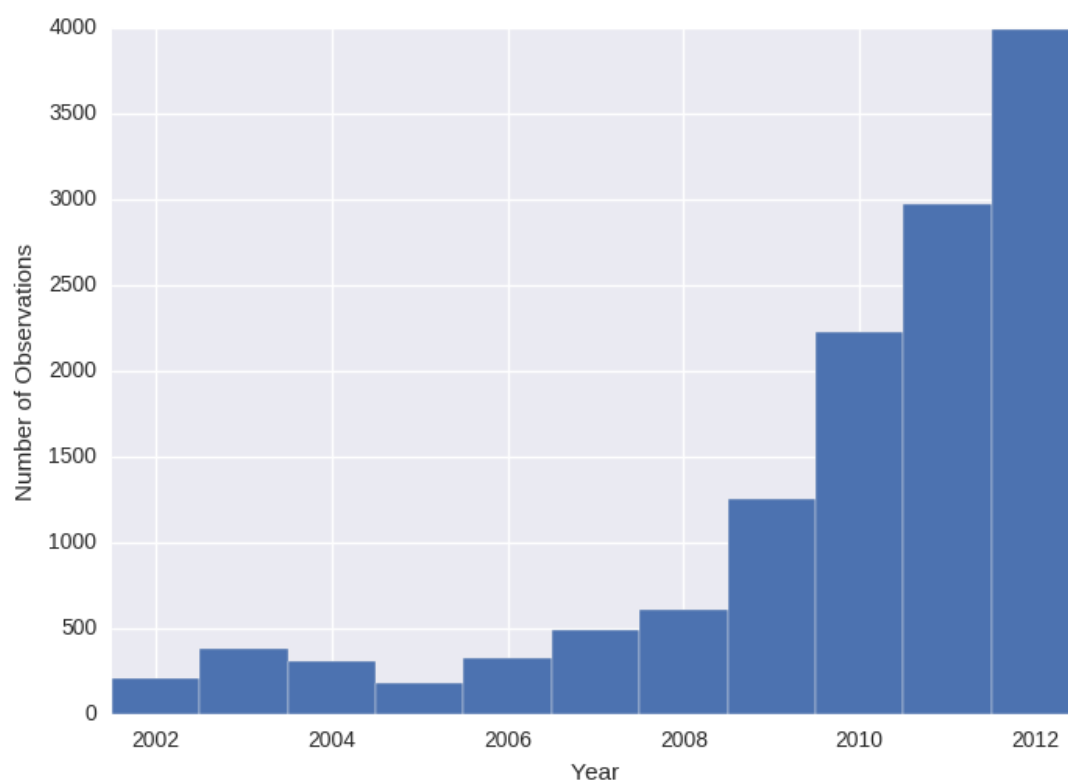


Figure 3.7. Temporal distribution of June observations in the ERD.

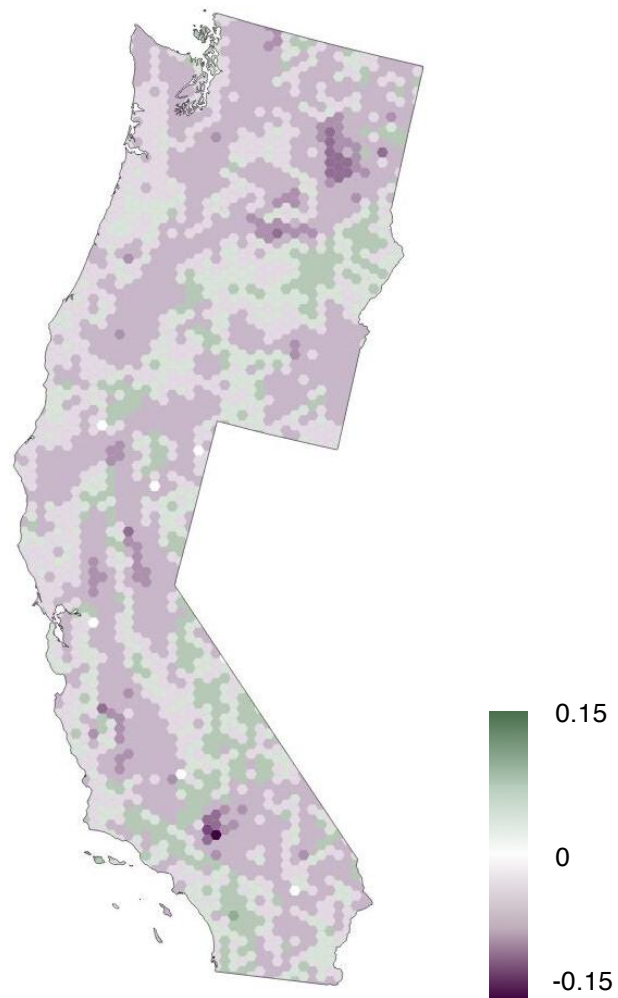


Figure 3.8. Difference of standard deviations between 1-year and 10-year Black-headed Grosbeak model. Standard deviations were calculated within 35,000 ha hexagons. Negative values (purple) occur where 1-year model standard deviations were higher.

Tables

Table 3.1. Predictors used for SDMs. Ecoregion and population density predictors were sourced from vector data, and thus, do not have a native resolution.

Variable	Source	Native resolution (m)
Elevation *	National Elevation Dataset	30
Ecoregion level III *	US EPA	N/A
Population density *	US Census Bureau	N/A
Mean annual precipitation *	PRISM	800
Mean annual min. temperature *	PRISM	800
Mean annual max. temperature *	PRISM	800
Year of observation	ERD	N/A
Canopy cover	LandTrendr	30
Impervious surface cover	LandTrendr	30
Land cover	LandTrendr	30
Mean canopy cover, 1.2 ha	LandTrendr	30
Mean impervious surface cover, 1.2 ha	LandTrendr	30
Percent bare ground, 1.2 ha	LandTrendr	30
Percent coniferous, 1.2 ha	LandTrendr	30
Percent deciduous, 1.2 ha	LandTrendr	30
Percent shrub, 1.2 ha	LandTrendr	30
Percent snow, 1.2 ha	LandTrendr	30
Percent water, 1.2 ha	LandTrendr	30
Mean canopy cover, 13.5 ha	LandTrendr	30
Mean impervious surface cover, 13.5 ha	LandTrendr	30
Percent bare ground, 13.5 ha	LandTrendr	30
Percent coniferous, 13.5 ha	LandTrendr	30
Percent deciduous, 13.5 ha	LandTrendr	30
Percent shrub, 13.5 ha	LandTrendr	30
Percent snow, 13.5 ha	LandTrendr	30
Percent water, 13.5 ha	LandTrendr	30

* Predictor that we did not adjust the temporal resolution for

CHAPTER 4: CONCLUSION

In this research, we have explored the effects of scale on distribution models by considering spatial and temporal scales important to two different phenomena across broad extents: land cover and bird species occurrence. While the mechanisms that affect each of these phenomena are quite different, the influence of scale is nonetheless an equally important consideration. Chapter 2 was primarily an investigation of spatial scale and how best to represent it in a broad-extent land cover model. The objectives of in that chapter were 1) to demonstrate the importance of representing local-scale patterns when making predictions for large areas, and 2) to use the models we developed to create yearly land cover time series that did not previously exist. Chapter 3 focused on the effects of temporal scale for modeling species distributions. Our primary objective in this chapter was to determine the appropriate temporal resolution(s) of land cover data for predicting species distributions. Although each chapter represented a discrete effort to explore the dynamics of scale, spatial patterns are closely related to temporal patterns (Adler and Lauenroth 2003; Ives and Klopfer 1997), and these two inter-related issues are therefore best considered together.

To examine spatial scale in Chapter 2, we applied an existing model algorithm, the spatiotemporal exploratory model (STEM) (Fink et al. 2010), in a novel context to develop yearly time series maps of forest canopy cover (FCC) and impervious surface cover (ISC). The STEM algorithm employs an ensemble of locally defined, spatially overlapping estimators to produce a map that reflects both local- and global-scale patterns across the modeling region. A model that remains sensitive to local-scale heterogeneity is vital for accurately predicting

distributions across broad extents because the variability in predictor-response relationships is likely to increase along with the extent of a modeling region. We ultimately used the FCC and ISC models to develop yearly time series maps of each, spanning 1990-2012.

To develop the most accurate time series of FCC and ISC, we first tested two model parameters, sample size and estimator size. Using the best-performing configuration of each, we then compared our STEMs to bagged decision trees, the most comparable model with globally defined estimators. We then used the best-performing STEM and LandTrendr, a novel time series of spectral imagery, to develop yearly maps of FCC and ISC, spanning 1990-2012. To test the temporal extensibility of our models, we compared our predicted 2011 maps to 2011 maps from the National Land Cover Database. Lastly, we developed several byproducts of our STEMs, variable importance maps, out-of-bag score maps, and standard deviation maps, which not only demonstrated the utility of the STEM methodology, but also provided important insights into predictor-response relationships of our models.

Our results comparing models trained on different numbers of samples showed an increase in performance with increasing sample size. We attributed the majority of performance gain to accuracy in predicting non-forest and non-impervious pixels. A training set will likely be adequate for developing a model if it sufficiently represents the variability of predictor-response relationships found throughout the entire modeling region. In Chapter 2, non-forest and non-impervious pixels represented a highly diverse class of materials representing a wide range of predictor-response patterns. A larger sample size, and greater number of non-forest and non-impervious samples in particular, was capable of adequately representing such diversity. These

results on sample size conflicted, however, with Coulston and others (2012), who found that a relatively small sample size was sufficient for making continental-scale FCC predictions. That study used a model with global estimators covering the entire modeling region, so more training samples could be shared across estimators. With the limited extent of estimators in a STEM, each estimator still required a sufficient number of samples to make predictions so more samples were necessary. This represents a considerable tradeoff with the STEM algorithm, as a large training set may not always be available for certain modeling tasks.

Since predictor-response patterns are likely to be more variable with a large modeling region, we expected to find decreasing model performance with increasing estimator size. We found this to be true for models of FCC and ISC. Estimator size is a reflection of the spatial scale important to the phenomenon modeled. These results demonstrated that predictor-response patterns changed over relatively short distances. Taking this comparison of estimator size further, the bagged decision tree models that we compared to STEMs were essentially a STEM, except that each estimator covered the entire extent of the modeling region. Bagged decision trees clearly performed worse than STEMs, as each estimator was forced to learn predictor-response patterns that were too complex to represent with a single estimator. Another way to interpret this is that larger estimators (and especially global estimators) did not properly reflect the spatial scale of predictor-response relationships for FCC and ISC.

While we used a STEM with a fixed estimator size, many processes may operate on multiple spatial scales (Swanson et al. 1988). A STEM that can represent multiple scales may be better suited to estimate distributions for phenomena that are influenced by such processes. In Chapter

2, we suggested that estimators defined by randomly delineated ecoregions could be one approach to achieve this, while also bounding the limits of estimators to by ecological gradients. A less computationally expensive method that might achieve a similar objective would be to randomly select the sizes of estimators from a range of heights and widths. This could allow a STEM to reflect processes at multiple scales.

The ultimate goal of developing accurate models of FCC and ISC predictor-response relationships was to create a time series of each, yearly datasets that did not previously exist at both a broad spatial extent and fine spatial and temporal resolution. To determine each models' ability to temporally extend those learned relationships to other years, we compared 2011 STEM predictions to 2011 predictions from the National Land Cover Database (NLCD), the only available land cover data at the same spatial resolution and broad extent. Although 2011 and 2001 ISC STEM and NLCD maps showed nearly equal agreement, the 2011 STEM FCC maps showed a marked drop in agreement. We showed, however, that this was likely the result of differences between the methods used to create 2001 and 2011 NLCD FCC maps.

This disparity in agreement also demonstrated differing qualities of maps produced when aggregating an ensemble by the mean or the mode. In developing our STEMs, we calculated both the mean and the mode of all spatially overlapping pixels to produce to different final maps for FCC and ISC each. We found that the maps created by the mode outperformed maps created by mean for all model iterations, however, 2011 NLCD FCC maps were developed with the mean of overlapping pixels. When we compared the 2011 STEM mean map with the 2011 NLCD FCC map, the two maps showed comparable agreement as the 2001 mode maps did. For

regression tasks where the final map is a continuous variable like FCC or ISC, the standard aggregation method is calculating the mean; the mode is more common for classification tasks, where predicted values represent discrete classes. Our results comparing the differences in mean versus mode predictions suggest that the mean might not always produce the most accurate final prediction. Since ISC and FCC generally represented as integers ranging from 0–100%, our use of the mode to aggregate estimations from an ensemble treated each integer value as a unique class. Whether this method ultimately produces more accurate predictions than the mean should be tested on an independent dataset, and it represents a topic for further study.

In our last investigation of spatial scale and its influences on FCC and ISC models, we developed three novel byproducts of our STEMs, all of which leverage the locally defined estimators that comprise a STEM: out-of-bag (OOB) score maps, standard deviation maps, and variable importance maps. OOB scores usually only yield a single evaluation score for an entire model, but with the each estimator of a STEM limited in spatial extent, we calculated an OOB score for each pixel. The result of this was a map estimating accuracy for each pixel of the map and that demonstrates the spatial variability in model performance. Similarly, we calculated the standard deviation for each pixel, which provided information on the uncertainty of final prediction values. Both of these maps generally showed error and uncertainty was higher in places where error and uncertainty in the source data was also higher. We created variable importance maps in much the same way as OOB score maps, except rather than calculating the OOB score for each pixel, we instead calculated the most important variable for determining that pixel's value in the final prediction map. Variable importance maps further supported the results of tests of estimator size, showing that importance varied across the modeling region for both FCC and ISC.

Furthermore, importance revealed patterns in predictor-response relationships that would have otherwise been obscured with an ensemble of globally defined estimators. Standard deviation, OOB score, and variable importance maps represent promising evaluation methods for assessing models of other phenomena and explaining local variation in predictor-response patterns.

In Chapter 3, we investigated the influence of temporal scale on an avian species distribution model (SDM) by comparing models developed with different temporal resolutions of predictor data. We were specifically interested in knowing how temporal resolution of land cover predictors would affect model performance. For fine-resolution models, the model training set contained bird observations with land cover predictor values that were contemporaneous with the observation. For coarse-resolution models, however, many training observations were related to land cover representing different years than the year the bird observation was made. Since different land cover types experience different rates of disturbance (Homer et al. 2015), we tested resolutions for SDMs of multiple species to see if habitat preference was related to SDM response to temporal resolution. We also calculated variable importance for each model to assess the structure and dynamics of models with differing resolutions.

We anticipated that coarse temporal resolution would generally produce less accurate models, but that SDM response would vary among species. Although SDMs for some species were less accurate at coarser resolutions, our results generally did not confirm this supposition. For many species, the best performing models were those produced at the coarsest resolution of 10 years. We suggested that the cause of this may be twofold: 1) time lags of species response to disturbances may have made coarse resolutions more reflective of ecological patterns important

to these species, and 2) coarser resolutions may have simplified models that were otherwise noisy at finer resolutions. Although some studies have shown that certain species may respond more than others to past landscape patterns (Ernoult et al. 2005; Knick and Rotenberry 2000), this could not alone explain our results. For every temporal resolution we tested, some observations were contemporaneous with land cover predictors, and at coarse resolutions, some observations were related to land cover representing more recent years. Thus, coarser temporal resolution did not always represent a time lag in the training data. Our other potential explanation, that coarse resolution models simplified predictor-response patterns, was supported by trends in variable importance for each model. With decreasing resolution, the importance of the year of the observation increased, suggesting that separating models by year was increasingly beneficial. Coarser resolution models, then, may have more easily been able to separate informative patterns (i.e., where the year of the observation was close to the year of land cover data) from uninformative ones (i.e., where the year of the observation did not match the year of land cover data). For fine-resolution models, there may have simply been too much variation in bird occurrence and land cover values to discern true signal from noise.

We also expected that species more commonly associated with habitats covering more inter-annually stable land cover types (e.g., anthropogenic landscapes such as urban and agricultural areas) would respond less to temporal resolution than species from land covers with higher frequencies of disturbance (e.g., forest). We found little evidence to support this proposition, although there were subtle indications that habitat preference may have influenced performance. For the two species associated with anthropogenic landscapes, Western Meadowlarks and Barn Swallows, the magnitude of difference between the best and worst performing models was lower

on average than two of the other species associated with forests, White-crowned Sparrows and Black-headed Grosbeaks. This effect was minor, however, and there were not any other results supporting the idea that habitat preference would influence SDM response to temporal resolution.

While we did not observe much difference in performance associated with habitat preference, it is possible that that other birds species may have been more responsive than those we selected. Modeling habitat for generalist species is usually more difficult than for specialists (Betts et al. 2006), and none of the species we modeled distributions for were true habitat specialists. Additionally, we used bird observations from eBird, a database of citizen science bird checklists (Dickinson et al. 2010). A commonly cited weakness of eBird is that bird observations are more likely to be collected in places where participants want to go (e.g., near popular birding hotspots, roads, and trails) rather than places that might offer a fuller picture of species occurrence (Hochachka and Fink 2012). As a result, the full effects of disturbances such as wildfire or clear-cutting may not be captured by these data. Whether or not the species or response data we selected in this study influenced our results are both topics for further study. Overall, our results from Chapter 3 demonstrate that temporal resolution can have a dramatic effect on SDM performance, but often in unexpected ways.

A common theme throughout both Chapters 2 and 3 was the importance of applying multiple evaluation metrics to assess model performance. A single metric can often provide an incomplete view of performance, as most metrics do not consider all possible types of error or they may be more influenced by certain types of error (Foody 2002; Ji and Gallo 2006; Riemann et al. 2010;

Strahler et al. 2006). In both chapters, we could only properly understand model performance by considering the results offered by multiple metrics. In Chapter 2, the kappa score decreased for ISC models with increasing estimator size, but the overall accuracy remained constant. This was largely because ISC models performed reasonably well when predicting for non-impervious pixels, and there are far more non-impervious pixels across the modeling region than impervious pixels. Kappa is essentially a ratio of the actual accuracy over the expected accuracy, which is generally higher with large map classes. Overall accuracy does not take into account the expected accuracy as kappa does, so the changes in performance were not reflected in just this one metric. Similarly, in Chapter 3, the OOB score showed nearly monotonic decreases performance with decreasing temporal resolution, regardless of how the models performed according to other metrics. This was likely because the OOB score was calculated with observations from the same geographic locations as the training samples. With spatial biases present in the eBird response data, the OOB score was more a measure of how well the models performed when predicting for locations similar to those typically visited by eBird participants. These results further confirm that multiple evaluation metrics are critical for any model evaluation, a practice that many authors have suggested become a regular part of model assessment (ibid).

Finally, the research presented in both Chapters 2 and 3 was only possible with source predictor data that were of fine resolution and broad extent—both spatial and temporal. The LandTrendr data we used here have opened up possibilities for exploring spatial and temporal dynamics of a range of environmental phenomena. LandTrendr represents only one of such ambitious efforts to develop spatiotemporal maps with broad applicability across disciplines (Hansen et al. 2013;

Huang et al. 2010; Kim et al. 2014, Townshend et al. 2012). In addition, recent improvements in computational efficiency including widespread cloud computing resources have made these initiatives even more accessible, enabling the processing of decades of global fine-resolution imagery in a figurative blink of the eye (Warren et al. 2015). In Chapter 2, we improved upon an existing algorithm for modeling distributions at a continental scale. This algorithm was computationally expensive, however, significantly impeding fast-paced model development. With cloud computing resources available to process in seconds what would normally take hours or days with conventional computing resources, the barrier of processing time and hardware cost no longer represents a significant impediment. The investigation of temporal resolution in Chapter 3 was also only possible with yearly time series data covering a broad extent. Publicly available platforms such as Google Earth Engine provide nearly complete multi-temporal datasets like the entire Landsat and MODIS archives, enabling unprecedented temporally explicit spatial analyses (Hansen et al. 2013; Pekel et al. 2016; Trianni et al. 2015). In concert with the methodological developments we have contributed and the considerations of temporal resolution that our research highlights, these novel computing platforms hold promise for powerful explorations of environmental systems and processes.

As these systems and processes continue to change as a result of global climate change, development pressure, and other deleterious impacts, understanding the distribution of environmental phenomena will be critical for effective conservation and management. The variability of predictor-response patterns for broad-extent models and nuanced SDM responses to temporal resolution that our research demonstrates are both illustrative of the complexity of

these phenomena. This research represents another step toward building tools and understanding for effective conservation of vital and potentially fragile systems.

References

- Adler, Peter B., and William K. Lauenroth. 2003. "The Power of Time: Spatiotemporal Scaling of Species Diversity." *Ecology Letters* 6 (8): 749–56. doi:10.1046/j.1461-0248.2003.00497.x.
- Betts, Matthew G., A.W. Diamond, G.J. Forbes, M.-A. Villard, and J.S. Gunn. 2006. "The Importance of Spatial Autocorrelation, Extent and Resolution in Predicting Forest Bird Occurrence." *Ecological Modelling* 191 (2): 197–224. doi:10.1016/j.ecolmodel.2005.04.027.
- Coulston, John W., Gretchen G. Moisen, Barry T. Wilson, Mark V. Finco, Warren B. Cohen, and C. Kenneth Brewer. 2012. "Modeling Percent Tree Canopy Cover: A Pilot Study." *Photogrammetric Engineering & Remote Sensing* 78 (7): 715–727.
- Dickinson, Janis L., Benjamin Zuckerberg, and David N. Bontar. 2010. "Citizen Science as an Ecological Research Tool: Challenges and Benefits." *Annual Review of Ecology, Evolution, and Systematics* 41 (1): 149–72. doi:10.1146/annurev-ecolsys-102209-144636.
- Ernault, Aude, Yves Tremauville, Dominique Cellier, Pierre Margerie, Estelle Langlois, and Didier Alard. 2006. "Potential Landscape Drivers of Biodiversity Components in a Flood Plain: Past or Present Patterns?" *Biological Conservation* 127 (1): 1–17. doi:10.1016/j.biocon.2005.07.008.
- Fink, Daniel, Wesley M. Hochachka, Benjamin Zuckerberg, David W. Winkler, Ben Shaby, M. Arthur Munson, Giles Hooker, Mirek Riedewald, Daniel Sheldon, and Steve Kelling. 2010. "Spatiotemporal Exploratory Models for Broad-Scale Survey Data." *Ecological Applications* 20 (8): 2131–2147.
- Foody, Giles M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80 (1): 185–201.
- Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342 (6160): 850–53. doi:10.1126/science.1244693.
- Hochachka, Wesley, and Daniel Fink. 2012. "Broad-Scale Citizen Science Data from Checklists: Prospects and Challenges for Macroecology." *Frontiers of Biogeography* 4 (4). <http://escholarship.org/uc/item/01t5c00w>.

- Homer, Collin, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Xian, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. 2015. "Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information." *Photogrammetric Engineering & Remote Sensing* 81 (5): 345–54. doi:10.1016/S0099-1112(15)30100-2.
- Huang, Chengquan, Samuel N. Goward, Jeffrey G. Masek, Nancy Thomas, Zhiliang Zhu, and James E. Vogelmann. 2010. "An Automated Approach for Reconstructing Recent Forest Disturbance History Using Dense Landsat Time Series Stacks." *Remote Sensing of Environment* 114 (1): 183–98. doi:10.1016/j.rse.2009.08.017.
- Ives, Anthony R., and Eric D. Klopfer. 1997. "Spatial Variation in Abundance Created by Stochastic Temporal Variation." *Ecology* 78 (6): 1907–1913.
- Ji, Lei, and Kevin Gallo. 2006. "An Agreement Coefficient for Image Comparison." *Photogrammetric Engineering & Remote Sensing* 72 (7): 823–833.
- Kim, Do-Hyung, Joseph O. Sexton, Praveen Noojipady, Chengquan Huang, Anupam Anand, Saurabh Channan, Min Feng, and John R. Townshend. 2014. "Global, Landsat-Based Forest-Cover Change from 1990 to 2000." *Remote Sensing of Environment* 155 (December): 178–93. doi:10.1016/j.rse.2014.08.017.
- Knick, Steven T., and John T. Rotenberry. 2000. "Ghosts of Habitats Past: Contribution of Landscape Change to Current Habitats Used by Shrubland Birds." *Ecology* 81 (1): 220–227.
- Pekel, Jean-François, Andrew Cottam, Noel Gorelick, and Alan S. Belward. 2016. "High-Resolution Mapping of Global Surface Water and Its Long-Term Changes." *Nature* 540 (7633): 418–22. doi:10.1038/nature20584.
- Riemann, Rachel, Barry Tyler Wilson, Andrew Lister, and Sarah Parks. 2010. "An Effective Assessment Protocol for Continuous Geospatial Datasets of Forest Characteristics Using USFS Forest Inventory and Analysis (FIA) Data." *Remote Sensing of Environment* 114 (10): 2337–52. doi:10.1016/j.rse.2010.05.010.
- Strahler, Alan H., Luigi Boschetti, Giles M. Foody, Mark A. Friedl, Matthew C. Hansen, Martin Herold, Philippe Mayaux, Jeffrey T. Morisette, Stephen V. Stehman, and Curtis E. Woodcock. 2006. "Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps." *European Communities, Luxembourg* 51. http://cndwebzine.hcp.ma/cnd_sii/IMG/pdf/Document2222222222-17.pdf.
- Swanson, F. J., T. K. Kratz, N. Caine, and R. G. Woodmansee. 1988. "Landform Effects on Ecosystem Patterns and Processes." *BioScience* 38 (2): 92–98. doi:10.2307/1310614.
- Townshend, John R., Jeffrey G. Masek, Chengquan Huang, Eric F. Vermote, Feng Gao, Saurabh Channan, Joseph O. Sexton, et al. 2012. "Global Characterization and Monitoring of Forest

Cover Using Landsat Data: Opportunities and Challenges.” *International Journal of Digital Earth* 5 (5): 373–97. doi:10.1080/17538947.2012.713190.

Trianni, Giovanna, Gianni Lisini, Emanuele Angiuli, E. A. Moreno, Piercarlo Dondi, Alessandro Gaggia, and Paolo Gamba. 2015. “Scaling up to National/Regional Urban Extent Mapping Using Landsat Data.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (7): 3710–19. doi:10.1109/JSTARS.2015.2398032.

Warren, Michael S., Steven P. Brumby, Samuel W. Skillman, Tim Kelton, Brendt Wohlberg, Mark Mathis, Rick Chartrand, Ryan Keisler, and Mark Johnson. 2015. “Seeing the Earth in the Cloud: Processing One Petabyte of Satellite Imagery in One Day.” In , 1–12. IEEE. doi:10.1109/AIPR.2015.7444536.

REFERENCES

- Abolafya, Moris, Ortaç Onmuş, Çağan H. Şekercioğlu, and Raşit Bilgin. 2013. "Using Citizen Science Data to Model the Distributions of Common Songbirds of Turkey Under Different Global Climatic Change Scenarios." Edited by Vanesa Magar. PLoS ONE 8 (7): e68037. doi:10.1371/journal.pone.0068037.
- Addicott, John F., John M. Aho, Michael F. Antolin, Dianna K. Padilla, John S. Richardson, and Daniel A. Soluk. 1987. "Ecological Neighborhoods: Scaling Environmental Patterns." *Oikos* 49 (3): 340–46. doi:10.2307/3565770.
- Adler, Peter B., and William K. Lauenroth. 2003. "The Power of Time: Spatiotemporal Scaling of Species Diversity." *Ecology Letters* 6 (8): 749–56. doi:10.1046/j.1461-0248.2003.00497.x.
- Araújo, Miguel B., and Mark New. 2007. "Ensemble Forecasting of Species Distributions." *Trends in Ecology & Evolution* 22 (1): 42–47. doi:10.1016/j.tree.2006.09.010.
- Arnold, Chester L., and C. James Gibbons. 1996. "Impervious Surface Coverage: The Emergence of a Key Environmental Indicator." *Journal of the American Planning Association* 62 (2): 243–58. doi:10.1080/01944369608975688.
- Ashbaugh, James G, and John A Alwin. 1994. *The Pacific Northwest: Geographical Perspectives*. Dubuque, Ia.: Kendall/Hunt Pub. Co.
- Ashcroft, Michael B., Kristine O. French, and Laurie A. Chisholm. 2011. "An Evaluation of Environmental Factors Affecting Species Distributions." *Ecological Modelling* 222 (3): 524–31. doi:10.1016/j.ecolmodel.2010.10.003.
- Austin, Mike P., and Kimberly P. Van Niel. 2011. "Improving Species Distribution Models for Climate Change Studies: Variable Selection and Scale." *Journal of Biogeography* 38 (1): 1–8. doi:10.1111/j.1365-2699.2010.02416.x.
- Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. "Aspect Transformation in Site Productivity Research." *Journal of Forestry* 64: 691–92.
- Betts, Matthew G., A.W. Diamond, G.J. Forbes, M.-A. Villard, and J.S. Gunn. 2006. "The Importance of Spatial Autocorrelation, Extent and Resolution in Predicting Forest Bird Occurrence." *Ecological Modelling* 191 (2): 197–224. doi:10.1016/j.ecolmodel.2005.04.027.
- Boakes, Elizabeth H., Philip J. K. McGowan, Richard A. Fuller, Ding Chang-qing, Natalie E. Clark, Kim O'Connor, and Georgina M. Mace. 2010. "Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data." *PLoS Biol* 8 (6): e1000385. doi:10.1371/journal.pbio.1000385.

- Bramer, Max. 2002. "Using J-Pruning to Reduce Overfitting of Classification Rules in Noisy Domains." In *International Conference on Database and Expert Systems Applications*, 433–442. Springer. http://link.springer.com/chapter/10.1007/3-540-46146-9_43.
- Breiman, Leo. 1984. *Classification and Regression Trees*. The Wadsworth Statistics/Probability Series. Belmont, Calif: Wadsworth International Group.
- . 1996a. "Bagging Predictors." *Machine Learning* 24 (2): 123–140.
- . 1996b. "Heuristics of Instability and Stabilization in Model Selection." *The Annals of Statistics* 24 (6): 2350–83.
- . 1998. "Arcing Classifiers." *The Annals of Statistics* 26 (3): 801–24.
- . 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.
- Budreski, Katherine, Michael Winchell, Lauren Padilla, JiSu Bang, and Richard A Brain. 2016. "A Probabilistic Approach for Estimating the Spatial Extent of Pesticide Agricultural Use Sites and Potential Co-Occurrence with Listed Species for Use in Ecological Risk Assessments: Probabilistic Crop Footprint Approach for Use in ESAs." *Integrated Environmental Assessment and Management* 12 (2): 315–27. doi:10.1002/ieam.1677.
- Bull, Joseph W, Kenwyn B Suttle, Navinder J Singh, and Ej Milner-Gulland. 2013. "Conservation When Nothing Stands Still: Moving Targets and Biodiversity Offsets." *Frontiers in Ecology and the Environment* 11 (4): 203–10. doi:10.1890/120020.
- Bulluck, Lesley, Erica Fleishman, Chris Betrus, and Rob Blair. 2006. "Spatial and Temporal Variations in Species Occurrence Rate Affect the Accuracy of Occurrence Models." *Global Ecology and Biogeography* 15 (1): 27–38. doi:10.1111/j.1466-822X.2006.00170.x.
- Campana, Néstor A, and Carlos E. M Tucci. 2001. "Predicting Floods from Urban Development Scenarios: Case Study of the Dilúvio Basin, Porto Alegre, Brazil." *Urban Water* 3 (1–2): 113–24. doi:10.1016/S1462-0758(01)00004-8.
- Carlson, Toby N, and S Traci Arthur. 2000. "The Impact of Land Use — Land Cover Changes due to Urbanization on Surface Microclimate and Hydrology: A Satellite Perspective." *Global and Planetary Change* 25 (1–2): 49–65. doi:10.1016/S0921-8181(00)00021-7.
- Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37–46. doi:10.1177/001316446002000104.
- Cohen, Warren B., Zhiqiang Yang, and Robert Kennedy. 2010. "Detecting Trends in Forest Disturbance and Recovery Using Yearly Landsat Time Series: 2. TimeSync — Tools for Calibration and Validation." *Remote Sensing of Environment* 114 (12): 2911–24. doi:10.1016/j.rse.2010.07.010.

- Cohen, Warren B., and Samuel N. Goward. 2004. "Landsat's Role in Ecological Applications of Remote Sensing." *BioScience* 54 (6): 535–45. doi:10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2.
- Coulston, John W., Gretchen G. Moisen, Barry T. Wilson, Mark V. Finco, Warren B. Cohen, and C. Kenneth Brewer. 2012. "Modeling Percent Tree Canopy Cover: A Pilot Study." *Photogrammetric Engineering & Remote Sensing* 78 (7): 715–727.
- Defries, R. S., M. C. Hansen, J. R. G. Townshend, A. C. Janetos, and T. R. Loveland. 2000. "A New Global 1-Km Dataset of Percentage Tree Cover Derived from Remote Sensing." *Global Change Biology* 6 (2): 247–54. doi:10.1046/j.1365-2486.2000.00296.x.
- Dickinson, Janis L., Benjamin Zuckerberg, and David N. Bontar. 2010. "Citizen Science as an Ecological Research Tool: Challenges and Benefits." *Annual Review of Ecology, Evolution, and Systematics* 41 (1): 149–72. doi:10.1146/annurev-ecolsys-102209-144636.
- Diefenbach, Duane R., Matthew R. Marshall, Jennifer A. Mattice, Daniel W. Brauning, and D. H. Johnson. 2007. "Incorporating Availability for Detection in Estimates of Bird Abundance." *The Auk* 124 (1): 96–106.
- Dietterich, Thomas G. 2000. "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization." *Machine Learning* 40 (2): 139–157.
- Elith, Jane, and John R. Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 677–97. doi:10.1146/annurev.ecolsys.110308.120159.
- Ernault, Aude, Yves Tremauville, Dominique Cellier, Pierre Margerie, Estelle Langlois, and Didier Alard. 2006. "Potential Landscape Drivers of Biodiversity Components in a Flood Plain: Past or Present Patterns?" *Biological Conservation* 127 (1): 1–17. doi:10.1016/j.biocon.2005.07.008.
- Ferraz, Katia Maria P. M. de Barros, Silvio Frosini de Barros Ferraz, Rogerio Cunha de Paula, Beatriz Beisiegel, and Christine Breitenmoser. 2012. "Species Distribution Modeling for Conservation Purposes." *Natureza & Conservação* 10 (2): 214–20. doi:10.4322/natcon.2012.032.
- Fink, Daniel, Theodoros Damoulas, Nicholas E. Bruns, Frank A. La Sorte, Wesley M. Hochachka, Carla P. Gomes, and Steve Kelling. 2014. "Crowdsourcing Meets Ecology: Hemisphere-Wide Spatiotemporal Species Distribution Models." *AI Magazine* 35 (2): 19–30.
- Fink, Daniel, Wesley M. Hochachka, Benjamin Zuckerberg, David W. Winkler, Ben Shaby, M. Arthur Munson, Giles Hooker, Mirek Riedewald, Daniel Sheldon, and Steve Kelling. 2010.

- “Spatiotemporal Exploratory Models for Broad-Scale Survey Data.” *Ecological Applications* 20 (8): 2131–2147.
- Foody, Giles M, and Peter M Atkinson. 2002. *Uncertainty in Remote Sensing and GIS*. Hoboken, NJ: J. Wiley. <http://public.eblib.com/choice/publicfullrecord.aspx?p=154955>.
- Foody, Giles M. 1992. “On the Compensation for Chance Agreement in Image Classification Accuracy Assessment.” *Photogrammetric Engineering And Remote Sensing* 58 (10): 1459–1460.
- . 2002. “Status of Land Cover Classification Accuracy Assessment.” *Remote Sensing of Environment* 80 (1): 185–201.
- . 2008. “Harshness in Image Classification Accuracy Assessment.” *International Journal of Remote Sensing* 29 (11): 3137–58. doi:10.1080/01431160701442120.
- Franklin, Janet, and Jennifer A Miller. 2009. *Mapping Species Distributions Spatial Inference and Prediction*. Cambridge; New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511810602>.
- Franklin, Jerry F., and C. T. Dyrness. 1988. *Natural Vegetation of Oregon and Washington*. Portland, Or., Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Dept. of Agriculture. <http://ir.library.oregonstate.edu/xmlui/handle/1957/34706>.
- Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. “The National Elevation Dataset.” *Photogrammetric Engineering and Remote Sensing* 68 (1): 5–+.
- Gottschalk, Thomas K., Birgit Aue, Stefan Hotes, and Klemens Ekschmitt. 2011. “Influence of Grain Size on Species–habitat Models.” *Ecological Modelling* 222 (18): 3403–12. doi:10.1016/j.ecolmodel.2011.07.008.
- Goward, Samuel N., Jeffrey G. Masek, Warren Cohen, Gretchen Moisen, G. James Collatz, Sean Healey, R. A. Houghton, et al. 2008. “Forest Disturbance and North American Carbon Flux.” *Eos, Transactions American Geophysical Union* 89 (11): 105–6. doi:10.1029/2008EO110001.
- Gray, Josh, and Conghe Song. 2013. “Consistent Classification of Image Time Series with Automatic Adaptive Signature Generalization.” *Remote Sensing of Environment* 134 (July): 333–41. doi:10.1016/j.rse.2013.03.022.
- Greenfield, Eric J., David J. Nowak, and Jeffrey T. Walton. 2009. “Assessment of 2001 NLCD Percent Tree and Impervious Cover Estimates.” *Photogrammetric Engineering & Remote Sensing* 75 (11): 1279–1286.
- Gschweng, Marion, Elisabeth K. V. Kalko, Peter Berthold, Wolfgang Fiedler, and Jakob Fahr. 2012. “Multi-Temporal Distribution Modelling with Satellite Tracking Data: Predicting

- Responses of a Long-Distance Migrant to Changing Environmental Conditions.” *Journal of Applied Ecology* 49 (4): 803–13. doi:10.1111/j.1365-2664.2012.02170.x.
- Gutzwiller, Kevin J., and Wylie C. Barrow. 2001. “Bird–Landscape Relations in the Chihuahuan Desert: Coping with Uncertainties About Predictive Models.” *Ecological Applications* 11 (5): 1517–32. doi:10.1890/1051-0761(2001)011[1517:BLRITC]2.0.CO;2.
- Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. “High-Resolution Global Maps of 21st-Century Forest Cover Change.” *Science* 342 (6160): 850–53. doi:10.1126/science.1244693.
- Hargrove, William W., and Forrest M. Hoffman. 2004. “Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions.” *Environmental Management* 34 (S1): S39–60. doi:10.1007/s00267-003-1084-0.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York.
<http://link.springer.com/10.1007/978-0-387-84858-7>.
- Hejl S. J., E. C. Beedy. 1986. “Weather induced variation in the abundance of birds.” In Verner, Jared, Michael L. Morrison, C. John Ralph, Pacific Southwest Forest and Range Experiment Station (Berkeley, Calif.), and Wildlife Society, eds. *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates: Based on an International Symposium Held at Stanford Sierra Camp, Fallen Leaf Lake, California, 7-11 October 1984*. Madison, Wis: University of Wisconsin Press, 1986.
- Ho, Tin Kam. 1998. “The Random Subspace Method for Constructing Decision Forests.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (8): 832–844.
- Hochachka, Wesley, and Daniel Fink. 2012. “Broad-Scale Citizen Science Data from Checklists: Prospects and Challenges for Macroecology.” *Frontiers of Biogeography* 4 (4).
<http://escholarship.org/uc/item/01t5c00w>.
- Homer, Collin, Chengquan Huang, Limin Yang, Bruce Wylie, and Michael Coan. 2004. “Development of a 2001 National Land-Cover Database for the United States.” *Photogrammetric Engineering & Remote Sensing* 70 (7): 829–840.
- Homer, Collin, Jon Dewitz, Joyce Fry, Michael Coan, Nazmul Hossain, Charles Larson, Nate Herold, et al. 2007. “Completion of the 2001 National Land Cover Database for the Conterminous United States.” *Photogrammetric Engineering and Remote Sensing* 73 (4): 337.
- Homer, Collin, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Xian, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. 2015. “Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing

- a Decade of Land Cover Change Information.” *Photogrammetric Engineering & Remote Sensing* 81 (5): 345–54. doi:10.1016/S0099-1112(15)30100-2.
- Hosmer, David W., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression*. Third edition. Wiley Series in Probability and Statistics. Hoboken, New Jersey: Wiley.
- Huang, C., L. S. Davis, and J. R. G. Townshend. 2002. “An Assessment of Support Vector Machines for Land Cover Classification.” *International Journal of Remote Sensing* 23 (4): 725–49. doi:10.1080/01431160110040323.
- Huang, Chengquan, Limin Yang, Bruce K. Wylie, and Collin Homer. 2001. “A Strategy for Estimating Tree Canopy Density Using Landsat 7 ETM+ and High Resolution Images over Large Areas.” <http://digitalcommons.unl.edu/usgspubs/111/>.
- Huang, Chengquan, Samuel N. Goward, Jeffrey G. Masek, Nancy Thomas, Zhiliang Zhu, and James E. Vogelmann. 2010. “An Automated Approach for Reconstructing Recent Forest Disturbance History Using Dense Landsat Time Series Stacks.” *Remote Sensing of Environment* 114 (1): 183–98. doi:10.1016/j.rse.2009.08.017.
- Ives, Anthony R., and Eric D. Klopfer. 1997. “Spatial Variation in Abundance Created by Stochastic Temporal Variation.” *Ecology* 78 (6): 1907–1913.
- Jepson, Wendy. 2005. “A Disappearing Biome? Reconsidering Land-Cover Change in the Brazilian Savanna.” *Geographical Journal* 171 (2): 99–111. doi:10.1111/j.1475-4959.2005.00153.x.
- Ji, Lei, and Kevin Gallo. 2006. “An Agreement Coefficient for Image Comparison.” *Photogrammetric Engineering & Remote Sensing* 72 (7): 823–833.
- Johnson, Douglas H. 2008. “In Defense of Indices: The Case of Bird Surveys.” *Journal of Wildlife Management* 72 (4): 857–68. doi:10.2193/2007-294.
- Johnson, Ned K. 1995. “Seven Avifaunal Censuses Spanning One-Half Century on an Island of White Firs (*Abies Concolor*) in the Mojave Desert.” *The Southwestern Naturalist* 40 (1): 76–85.
- Johnson, Roger W. 2001. “An Introduction to the Bootstrap.” *Teaching Statistics* 23 (2): 49.
- Johnston, Alison, Daniel Fink, Mark D. Reynolds, Wesley M. Hochachka, Brian L. Sullivan, Nicholas E. Bruns, Eric Hallstein, Matt S. Merrifield, Sandi Matsumoto, and Steve Kelling. 2015. “Abundance Models Improve Spatial and Temporal Prioritization of Conservation Resources.” *Ecological Applications* 25 (7): 1749–1756.
- Kennedy, Robert E., Zhiqiang Yang, Justin Braaten, Catharine Copass, Natalya Antonova, Chris Jordan, and Peder Nelson. 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 (September): 271–85. doi:10.1016/j.rse.2015.05.005.
- Kennedy, Robert E., Zhiqiang Yang, and Warren B. Cohen. 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. doi:10.1016/j.rse.2010.07.008.
- Kennedy, Robert E., Zhiqiang Yang, Warren B. Cohen, Eric Pfaff, Justin Braaten, and Peder Nelson. 2012. “Spatial and Temporal Patterns of Forest Disturbance and Regrowth within the Area of the Northwest Forest Plan.” *Remote Sensing of Environment* 122 (July): 117–33. doi:10.1016/j.rse.2011.09.024.
- Kim, Do-Hyung, Joseph O. Sexton, Praveen Noojipady, Chengquan Huang, Anupam Anand, Saurabh Channan, Min Feng, and John R. Townshend. 2014. “Global, Landsat-Based Forest-Cover Change from 1990 to 2000.” *Remote Sensing of Environment* 155 (December): 178–93. doi:10.1016/j.rse.2014.08.017.
- Knick, Steven T., and John T. Rotenberry. 2000. “Ghosts of Habitats Past: Contribution of Landscape Change to Current Habitats Used by Shrubland Birds.” *Ecology* 81 (1): 220–227.
- Liu, Canran, Paul Frazier, and Lalit Kumar. 2007. “Comparative Assessment of the Measures of Thematic Classification Accuracy.” *Remote Sensing of Environment* 107 (4): 606–16. doi:10.1016/j.rse.2006.10.010.
- Lunetta, Ross S, David M Johnson, John G Lyon, and Jill Crotnell. 2004. “Impacts of Imagery Temporal Frequency on Land-Cover Change Detection Monitoring.” *Remote Sensing of Environment* 89 (4): 444–54. doi:10.1016/j.rse.2003.10.022.
- MacEachren, Alan M. 1992. “Visualizing Uncertain Information.” *Cartographic Perspectives*, no. 13: 10–19.
- Manel, S., S.t. Buckton, and S.j. Ormerod. 2000. “Testing Large-Scale Hypotheses Using Surveys: The Effects of Land Use on the Habitats, Invertebrates and Birds of Himalayan Rivers.” *Journal of Applied Ecology* 37 (5): 756–70. doi:10.1046/j.1365-2664.2000.00537.x.
- Martinuzzi, Sebastia’n. 2009. “Incorporating Remotely Sensed Tree Canopy Cover Data into Broad Scale Assessments of Wildlife Habitat Distribution and Conservation.” *Journal of Applied Remote Sensing* 3 (1): 033568. doi:10.1117/1.3279080.
- Masek, Jeffrey G., Chengquan Huang, Robert Wolfe, Warren Cohen, Forrest Hall, Jonathan Kutler, and Peder Nelson. 2008. “North American Forest Disturbance Mapped from a Decadal Landsat Record.” *Remote Sensing of Environment* 112 (6): 2914–26. doi:10.1016/j.rse.2008.02.010.

- McClintock, Brett T., Larissa L. Bailey, Kenneth H. Pollock, and Theodore R. Simons. 2010. "Unmodeled Observation Error Induces Bias When Inferring Patterns and Dynamics of Species Occurrence via Aural Detections." *Ecology* 91 (8): 2446–2454.
- Miller, Crane S., and Richard S. Hyslop. 1983. *California, the Geography of Diversity*. 1st ed. Palo Alto, Calif: Mayfield Pub. Co.
- Mitchell, Michael S., Richard A. Lancia, and John A. Gerwin. 2001. "Using Landscape-Level Data to Predict the Distribution of Birds on a Managed Forest: Effects of Scale." *Ecological Applications* 11 (6): 1692–1708. doi:10.1890/1051-0761(2001)011[1692:ULLDTP]2.0.CO;2.
- Munson, M. Arthur, Kevin Webb, Daniel Sheldon, Daniel Fink, Wesley M. Hochachka, Marshall Iliff, Mirek Riedewald, Daria Sorokina, Brian Sullivan, Christopher Wood, and Steve Kelling. "The eBird Reference Dataset, Version 5.0." Cornell Lab of Ornithology and National Audubon Society, Ithaca, NY, January 2013.
- Murphy, Helen T., and Jon Lovett-Doust. 2007. "Accounting for Regional Niche Variation in Habitat Suitability Models." *Oikos* 116 (1): 99–110. doi:10.1111/j.2006.0030-1299.15050.x.
- Nagy, Christopher, Kyle Bardwell, Robert F. Rockwell, Rod Christie, and Mark Weckel. 2012. "Validation of a Citizen Science-Based Model of Site Occupancy for Eastern Screech Owls with Systematic Data in Suburban New York and Connecticut." <https://www.bioone.org/doi/full/10.1656/045.019.s611>.
- Nemani, Ramakrishna R., and Steven W. Running. 1995. "Satellite Monitoring of Global Land Cover Changes and Their Impact on Climate." *Climatic Change* 31 (2–4): 395–413. doi:10.1007/BF01095154.
- Nowak, David J., and Eric J. Greenfield. 2010. "Evaluating The National Land Cover Database Tree Canopy and Impervious Cover Estimates Across the Conterminous United States: A Comparison with Photo-Interpreted Estimates." *Environmental Management* 46 (3): 378–90. doi:10.1007/s00267-010-9536-9.
- Olofsson, Pontus, Giles M. Foody, Stephen V. Stehman, and Curtis E. Woodcock. 2013. "Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty Using Stratified Estimation." *Remote Sensing of Environment* 129 (February): 122–31. doi:10.1016/j.rse.2012.10.031.
- Opitz, David, and Richard Maclin. 1999. "Popular Ensemble Methods: An Empirical Study." *Journal of Artificial Intelligence Research*, 169–198.
- PRISM Climate Group. 2012. "30-Year Normals." Oregon State University. Corvallis, OR. <http://prism.oregonstate.edu>. Version 14.1-20140502-1000. Created 11 Jul 2012.

- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (Oct): 2825–2830.
- Pekel, Jean-François, Andrew Cottam, Noel Gorelick, and Alan S. Belward. 2016. "High-Resolution Mapping of Global Surface Water and Its Long-Term Changes." *Nature* 540 (7633): 418–22. doi:10.1038/nature20584.
- Peterson, A. Townsend, and Robert D. Holt. 2003. "Niche Differentiation in Mexican Birds: Using Point Occurrences to Detect Ecological Innovation." *Ecology Letters* 6 (8): 774–82. doi:10.1046/j.1461-0248.2003.00502.x.
- Pidgeon, Anna M., Volker C. Radeloff, Curt H. Flather, Christopher A. Lepczyk, Murray K. Clayton, Todd Jerome Hawbaker, and Roger B. Hammer. 2007. "Associations of Forest Bird Species Richness with Housing and Landscape Patterns across the USA." *Ecological Applications* 17 (7): 1989–2010.
- Pielke, Roger A., Andy Pitman, Dev Niyogi, Rezaul Mahmood, Clive McAlpine, Faisal Hossain, Kees Klein Goldewijk, et al. 2011. "Land Use/Land Cover Changes and Climate: Modeling Analysis and Observational Evidence." *Wiley Interdisciplinary Reviews: Climate Change* 2 (6): 828–50. doi:10.1002/wcc.144.
- Pontius, Robert Gilmore, and Marco Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing* 32 (15): 4407–29. doi:10.1080/01431161.2011.552923.
- Quevedo, Mario, Richard Svanbäck, and Peter Eklöv. 2009. "Intrapopulation Niche Partitioning in a Generalist Predator Limits Food Web Connectivity." *Ecology* 90 (8): 2263–2274.
- Radchuk, Viktoriia, Karin Johst, Jürgen Groeneveld, Camille Turlure, Volker Grimm, and Nicolas Schtickzelle. 2014. "Appropriate Resolution in Time and Model Structure for Population Viability Analysis: Insights from a Butterfly Metapopulation." *Biological Conservation* 169 (January): 345–54. doi:10.1016/j.biocon.2013.12.004.
- Riemann, Rachel, Barry Tyler Wilson, Andrew Lister, and Sarah Parks. 2010. "An Effective Assessment Protocol for Continuous Geospatial Datasets of Forest Characteristics Using USFS Forest Inventory and Analysis (FIA) Data." *Remote Sensing of Environment* 114 (10): 2337–52. doi:10.1016/j.rse.2010.05.010.
- Rullman, Stan, and John M. Marzluff. 2014. "Raptor Presence Along an Urban–Wildland Gradient: Influences of Prey Abundance and Land Cover." *Journal of Raptor Research* 48 (3): 257–72. doi:10.3356/JRR-13-32.1.
- Sanyal, J, Song Zhang, J Dyer, A Mercer, P Amburn, and R J Moorhead. 2010. "Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty." *IEEE*

- Transactions on Visualization and Computer Graphics 16 (6): 1421–30.
doi:10.1109/TVCG.2010.181.
- Schimel, David S., Vemap Participants, and B. H. Braswell. 1997. “Continental Scale Variability in Ecosystem Processes: Models, Data, and the Role of Disturbance.” *Ecological Monographs* 67 (2): 251–71. doi:10.1890/0012-9615(1997)067[0251:CSVIEP]2.0.CO;2.
- Schroeder, Todd A., Warren B. Cohen, and Zhiqiang Yang. 2007. “Patterns of Forest Regrowth Following Clearcutting in Western Oregon as Determined from a Landsat Time-Series.” *Forest Ecology and Management* 243 (2–3): 259–73. doi:10.1016/j.foreco.2007.03.019.
- Scott, J. Michael, ed. 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Washington, DC: Island Press.
- Seo, C., J. H Thorne, L. Hannah, and W. Thuiller. 2009. “Scale Effects in Species Distribution Models: Implications for Conservation Planning under Climate Change.” *Biology Letters* 5 (1): 39–43. doi:10.1098/rsbl.2008.0476.
- Shao, Yang, and Ross S. Lunetta. 2012. “Comparison of Support Vector Machine, Neural Network, and CART Algorithms for the Land-Cover Classification Using Limited Training Data Points.” *ISPRS Journal of Photogrammetry and Remote Sensing* 70 (June): 78–87. doi:10.1016/j.isprsjprs.2012.04.001.
- Sheeder, Scott A., Jeremy D. Ross, and Toby N. Carlson. 2002. “Dual Urban and Rural Hydrograph Signals in Three Small Watersheds1.” *JAWRA Journal of the American Water Resources Association* 38 (4): 1027–40. doi:10.1111/j.1752-1688.2002.tb05543.x.
- Sheeren, D., S. Bonthoux, and G. Balent. 2014. “Modeling Bird Communities Using Unclassified Remote Sensing Imagery: Effects of the Spatial Resolution and Data Period.” *Ecological Indicators* 43 (August): 69–82. doi:10.1016/j.ecolind.2014.02.023.
- Simon, N Pp, F E Schwab, and R D Otto. 2002. “Songbird Abundance in Clear-Cut and Burned Stands: A Comparison of Natural Disturbance and Forest Management.” *Canadian Journal of Forest Research* 32 (8): 1343–50. doi:10.1139/x02-057.
- Song, Wonkyong, Eunyoung Kim, Dongkun Lee, Mounjin Lee, and Seong-Woo Jeon. 2013. “The Sensitivity of Species Distribution Modeling to Scale Differences.” *Ecological Modelling* 248 (January): 113–18. doi:10.1016/j.ecolmodel.2012.09.012.
- Stephens, Jaime L., and John D. Alexander. 2011. “Effects of Fuel Reduction on Bird Density and Reproductive Success in Riparian Areas of Mixed-Conifer Forest in Southwest Oregon.” *Forest Ecology and Management* 261 (1): 43–49. doi:10.1016/j.foreco.2010.09.022.
- Strahler, Alan H., Luigi Boschetti, Giles M. Foody, Mark A. Friedl, Matthew C. Hansen, Martin Herold, Philippe Mayaux, Jeffrey T. Morisette, Stephen V. Stehman, and Curtis E.

- Woodcock. 2006. "Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps." European Communities, Luxembourg 51. http://cndwebzine.hcp.ma/cnd_sii/IMG/pdf/Document2222222222-17.pdf.
- Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution." *BMC Bioinformatics* 8: 25. doi:10.1186/1471-2105-8-25.
- Suriya, S., and B. V. Mudgal. 2012. "Impact of Urbanization on Flooding: The Thirusoolam Sub Watershed – A Case Study." *Journal of Hydrology, Hydrology Conference 2010*, 412–413 (January): 210–19. doi:10.1016/j.jhydrol.2011.05.008.
- Swanson, F. J., T. K. Kratz, N. Caine, and R. G. Woodmansee. 1988. "Landform Effects on Ecosystem Patterns and Processes." *BioScience* 38 (2): 92–98. doi:10.2307/1310614.
- Theobald, David M., Scott J. Goetz, John B. Norman, and Patrick Jantz. 2009. "Watersheds at Risk to Increased Impervious Surface Cover in the Conterminous United States." *Journal of Hydrologic Engineering* 14 (4): 362–68. doi:10.1061/(ASCE)1084-0699(2009)14:4(362).
- Townshend, John R., Jeffrey G. Masek, Chengquan Huang, Eric F. Vermote, Feng Gao, Saurabh Channan, Joseph O. Sexton, et al. 2012. "Global Characterization and Monitoring of Forest Cover Using Landsat Data: Opportunities and Challenges." *International Journal of Digital Earth* 5 (5): 373–97. doi:10.1080/17538947.2012.713190.
- Trianni, Giovanna, Gianni Lisini, Emanuele Angiuli, E. A. Moreno, Piercarlo Dondi, Alessandro Gaggia, and Paolo Gamba. 2015. "Scaling up to National/Regional Urban Extent Mapping Using Landsat Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (7): 3710–19. doi:10.1109/JSTARS.2015.2398032.
- Turner, D. P., W. D. Ritts, B. E. Law, W. B. Cohen, Z. Yang, T. Hudiburg, J. L. Campbell, and M. Duane. 2007. "Scaling Net Ecosystem Production and Net Biome Production over a Heterogeneous Region in the Western United States." *Biogeosciences Discussions* 4 (2): 1093–1135.
- US Census Bureau. 2013. "2010 Census Urban Area Facts." March 7. <http://www.census.gov/geo/reference/ua/uafacts.html>.
- US Environmental Protection Agency. 2012. "U.S. Level III and IV Ecoregions (U.S. EPA)." Research Triangle Park, NC. <https://catalog.data.gov/dataset/u-s-level-iii-and-iv-ecoregions-u-s-epa>. Created 26 Apr 2012.
- Vaughan, I. P., and S. J. Ormerod. 2003. "Improving the Quality of Distribution Models for Conservation by Addressing Shortcomings in the Field Collection of Training Data." *Conservation Biology* 17 (6): 1601–11. doi:10.1111/j.1523-1739.2003.00359.x.

- Vierling, Kerri T., Charles E. Swift, Andrew T. Hudak, Jody C. Vogeler, and Lee A. Vierling. 2014. "How Much Does the Time Lag between Wildlife Field-Data Collection and LiDAR-Data Acquisition Matter for Studies of Animal Distributions? A Case Study Using Bird Communities." *Remote Sensing Letters* 5 (2): 185–93. doi:10.1080/2150704X.2014.891773.
- Wang, Jie, Yuanyuan Zhao, Congcong Li, Le Yu, Desheng Liu, and Peng Gong. 2015. "Mapping Global Land Cover in 2001 and 2010 with Spatial-Temporal Consistency at 250m Resolution." *ISPRS Journal of Photogrammetry and Remote Sensing* 103 (May): 38–47. doi:10.1016/j.isprsjprs.2014.03.007.
- Warren, Michael S., Steven P. Brumby, Samuel W. Skillman, Tim Kelton, Brendt Wohlberg, Mark Mathis, Rick Chartrand, Ryan Keisler, and Mark Johnson. 2015. "Seeing the Earth in the Cloud: Processing One Petabyte of Satellite Imagery in One Day." In , 1–12. IEEE. doi:10.1109/AIPR.2015.7444536.
- White, J. C., M. A. Wulder, G. W. Hobart, J. E. Luther, T. Hermosilla, P. Griffiths, N. C. Coops, et al. 2014. "Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science." *Canadian Journal of Remote Sensing* 40 (3): 192–212. doi:10.1080/07038992.2014.945827.
- Wickham, J.D., T.G. Wade, and D.J. Norton. 2014. "Spatial Patterns of Watershed Impervious Cover Relative to Stream Location." *Ecological Indicators* 40 (May): 109–16. doi:10.1016/j.ecolind.2014.01.013.
- Wilson, B. Tyler, Andrew J. Lister, and Rachel I. Riemann. 2012. "A Nearest-Neighbor Imputation Approach to Mapping Tree Species over Large Areas Using Forest Inventory Plots and Moderate Resolution Raster Data." *Forest Ecology and Management* 271 (May): 182–98. doi:10.1016/j.foreco.2012.02.002.
- Xin, Qinchuan, Pontus Olofsson, Zhe Zhu, Bin Tan, and Curtis E. Woodcock. 2013. "Toward near Real-Time Monitoring of Forest Disturbance by Fusion of MODIS and Landsat Data." *Remote Sensing of Environment* 135 (August): 234–47. doi:10.1016/j.rse.2013.04.002.
- Yu, Jun, Weng-Keen Wong, and Steve Kelling. 2014. "Clustering Species Accumulation Curves to Identify Skill Levels of Citizen Scientists Participating in the eBird Project." <http://web.engr.oregonstate.edu/~wong/papers/pdf/IAAI2014.SAC.pdf>.
- Zhu, Chunhao, and Yingkui Li. 2013. "High Temporal Resolution Land Use/Land Cover Change from 1984 to 2010 of the Little River Watershed, Tennessee, Investigated Using Landsat and Google Earth Images." *Southeastern Geographer* 53 (3): 250–266.

APPENDIX A: CONFUSION MATRICES FOR STEM MODELS WITH MEDIUM AND LARGE SUPPORT SETS

Table A.1. Confusion matrix for medium-sized (400 km x 300 km) forest canopy cover model.

		Reference map class										Total ref.	User's (%)	Area (ha)
		0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100			
Predicted map class	0–10	0.623	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.002	0	0.641	97.2	604497242
	11–20	0.007	0.013	0.006	0.001	0	0	0	0	0	0	0.028	46.9	20993343
	21–30	0.002	0.001	0.018	0.005	0.002	0	0	0	0	0	0.029	61.6	25816593
	31–40	0	0	0.002	0.020	0.005	0.001	0	0	0	0	0.030	66.3	27564027
	41–50	0	0	0	0.003	0.023	0.005	0.001	0	0	0	0.034	68.8	30905594
	51–60	0	0	0	0	0.004	0.026	0.005	0	0	0	0.037	70.2	33429829
	61–70	0	0	0	0	0	0.003	0.029	0.006	0	0	0.040	72.4	35884371
	71–80	0	0	0	0	0	0	0.003	0.035	0.006	0	0.045	77.4	41327400
	81–90	0	0	0	0	0	0	0	0.003	0.061	0.005	0.070	87.4	66497342
	91–100	0	0	0	0	0	0	0	0	0.004	0.042	0.046	91.2	42816721
Total pred.		0.634	0.019	0.030	0.033	0.036	0.038	0.040	0.047	0.075	0.049			
Producer's (%)		98.3	70.8	61.4	61.8	64.2	68.5	70.7	73.4	81.3	86.5			

Table A.2. Confusion matrix for large-sized (800 km x 600 km) forest canopy cover model.

		Reference map class										Total ref.	User's (%)	Area (ha)
		0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100			
Predicted map class	0–10	0.626	0.004	0.004	0.003	0.002	0.002	0.001	0.001	0.003	0	0.648	96.7	612258975
	11–20	0.007	0.012	0.005	0.001	0	0	0	0	0	0	0.026	45.7	19442378
	21–30	0.002	0.001	0.016	0.005	0.002	0	0	0	0	0	0.028	59.4	24038798
	31–40	0	0	0.002	0.019	0.005	0.001	0	0	0	0	0.030	63.2	26641111
	41–50	0	0	0	0.003	0.022	0.005	0.001	0	0	0	0.033	67.7	30512251
	51–60	0	0	0	0	0.004	0.025	0.005	0.001	0	0	0.037	68.2	33362898
	61–70	0	0	0	0	0	0.003	0.027	0.006	0.001	0	0.039	70.9	35053924
	71–80	0	0	0	0	0	0	0.003	0.033	0.006	0	0.043	76.1	39641954
	81–90	0	0	0	0	0	0	0	0.004	0.058	0.005	0.068	85.5	64826373
	91–100	0	0	0	0	0	0	0	0	0.005	0.042	0.047	87.7	43953800
Total pred.		0.637	0.018	0.028	0.032	0.037	0.038	0.040	0.046	0.074	0.048			
Producer's (%)		98.3	66.8	58.7	59.4	61.2	66.0	68.5	70.7	78.2	85.6			

Table A.3. Confusion matrix for medium-sized (400 km x 300 km) impervious surface cover model

		Reference map class										Total ref.	User's (%)	Area (ha)
		0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100			
Predicted map class	0–10	0.9356	0.0025	0.0008	0.0004	0.0002	0.0001	0	0	0	0	0.940	99.6	902759571
	11–20	0.0041	0.0028	0.0006	0	0	0	0	0	0	0	0.008	36.2	4359654
	21–30	0.0016	0.0007	0.0028	0.0006	0	0	0	0	0	0	0.006	48.7	4189836
	31–40	0.0007	0	0.0007	0.0027	0.0006	0	0	0	0	0	0.005	55.8	3985891
	41–50	0.0004	0	0	0.0005	0.0031	0.0005	0	0	0	0	0.005	64.9	4090046
	51–60	0.0003	0	0	0	0.0005	0.0031	0.0004	0	0	0	0.004	70.4	3815831
	61–70	0.0002	0	0	0	0	0.0004	0.0022	0.0003	0	0	0.003	72.0	2745695
	71–80	0.0001	0	0	0	0	0	0.0002	0.0015	0.0002	0	0.002	71.3	1799112
	81–90	0	0	0	0	0	0	0	0.0001	0.0011	0	0.001	78.8	1236370
	91–100	0	0	0	0	0	0	0	0	0	0.0006	0.001	84.9	580950
Total pred.		0.943	0.006	0.005	0.004	0.005	0.004	0.003	0.002	0.001	0.001			
Producer's (%)		99.2	45.7	56.3	63.0	68.2	72.1	73.0	74.3	78.8	85.0			

Table A.4. Confusion matrix for large-sized (800 km x 600 km) impervious surface cover model.

		Reference map class										Total ref.	User's (%)	Area (ha)
		0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100			
Predicted map class	0–10	0.9305	0.0027	0.0010	0.0004	0.0003	0.0002	0	0	0	0	0.935	99.5	903933497
	11–20	0.0034	0.0022	0.0005	0	0	0	0	0	0	0	0.006	35.6	3515954
	21–30	0.0013	0.0006	0.0026	0.0005	0	0	0	0	0	0	0.005	50.3	3861806
	31–40	0.0007	0	0.0007	0.0026	0.0006	0	0	0	0	0	0.005	55.0	3835366
	41–50	0.0004	0	0	0.0005	0.0031	0.0005	0	0	0	0	0.005	65.3	4076928
	51–60	0.0003	0	0	0	0.0005	0.0031	0.0004	0	0	0	0.004	71.1	3902504
	61–70	0.0002	0	0	0	0	0.0004	0.0022	0.0003	0	0	0.003	71.2	2739792
	71–80	0.0002	0	0	0	0	0	0.0002	0.0015	0.0002	0	0.002	70.4	1821383
	81–90	0	0	0	0	0	0	0	0.0001	0.0011	0	0.001	75.7	1260175
	91–100	0	0	0	0	0	0	0	0	0	0.0006	0.001	83.4	615551
Total pred.		0.937	0.006	0.005	0.004	0.005	0.004	0.003	0.002	0.001	0.001			
Producer's (%)		99.3	39.6	53.6	61.6	67.6	71.7	71.7	73.8	78.2	84.8			

