US Department of Agriculture Forest Service
Pacific Northwest Research Station
Northern Spotted Owl Passive Acoustic Monitoring
2018 Annual Report
June, 2019

1. Title

Research update on using passive acoustics within a random sampling framework to monitor northern spotted owl populations in Washington and Oregon.

2. Research Team

Damon B. Lesmeister[1], Raymond J. Davis[2], Leila S. Duchac[3], and Zachary J. Ruff[4]

[1]Research Wildlife Biologist, USDA Forest Service, Pacific Northwest Research Station *and* Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR.
[2]Monitoring Lead - Older Forests and Spotted Owls, USDA Forest Service, Pacific Northwest Region, Corvallis, OR.
[3]Graduate Student, USDA Forest Service, Pacific Northwest Research Station *and* Oregon Cooperative Fish and Wildlife Research Unit, Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR.
[4]Research Assistant, USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR.

3. Introduction

As part of the Effectiveness Monitoring Program for the Northwest Forest Plan, northern spotted owl (*Strix occidentalis caurina*; hereafter NSO) demography has been monitored for over two decades with mark-resighting methods on 8 study areas comprised primarily of federal lands in the Pacific Northwest (Lint et al. 1999). In addition to capturing, marking, and resighting birds, the study design entails callback surveys to locate territorial owls (Dugger et al. 2016). This has proven to be highly effective at estimating trends in basic vital rates like survival and reproductive success and the annual rate of population change throughout the subspecies' geographic range, as well as identifying factors associated with those trends (e.g., Anthony et al. 2006, Forsman et al. 2011, Dugger et al. 2016). However, the increasing presence of barred owls (S. varia) and declining presence of NSO have greatly increased the amount of effort and costs it takes to accomplish these studies. Furthermore, several NSO populations, especially in Washington, have declined to levels where few individuals occupy and reproduce in the historic monitoring territories, thus increasing uncertainty in population status and trend estimates. For example, the Olympic Peninsula population estimate for 2011 was about 40% (as low as possibly 10%) of the initial population estimate for 1992 (Dugger et al. 2016). As with all known populations across the subspecies range, the population on the Olympic Peninsula has continued to decline since 2011(Gremel 2019, Lesmeister et al. 2019); bringing into question the sustainability of long-term mark-recapture methods for estimating demography on this study area. Beginning in 2017, we began a pilot test using passive bioacoustics recorders for a random census design (i.e., occupancy-based framework), as considered in the NSO Effectiveness

Monitoring Program, but without vocal call lures (Lint et al. 1999). In 2018 the decision was made by the Regional Interagency Executive Committee to discontinue mark-resighting methods to monitor NSOs on the Olympic Peninsula, and beginning in 2019 to use passive bioacoustic monitoring for that population.

Aside from the increased uncertainty associated with estimates of vital rates when NSO populations reach low levels, there are several general short- and long-term drawbacks to a key aspect of mark-recapture methods, which are callback surveys designed to locate territorial individuals. These surveys can cause behavioral changes that can negatively affect an individual's overall fitness as well as decreasing detectability (Conway and Gibbs 2005), concerns that may be especially relevant when surveying for threatened and endangered species. Specific concerns are: 1) callback surveys elicit a territorial response of a specific species, therefore, males are more likely to be detected than females, as was found in burrowing and tawny owls (Moulton et al. 2004, Worthington-Hill and Conway 2017); 2) individuals change their behavior in response to broadcast conspecific calls, often by leaving their nest sites to approach the broadcast source, exposing themselves and the nest to predation (Haug and Didiuk 1993, Langham et al. 2006, Conway et al. 2008); 3) females may perceive conspecific broadcast calls as a male in a song contest with her mate, and when the recorded song was more aggressive (longer and louder) than that of the female's mate, the male suffered decreased fitness as his mate was much more likely to seek extra-pair copulations (Mennill et al. 2002); 4) male birds can experience a spike in testosterone in response to broadcast calls, with the increase more marked in species with monogamous mating systems (Wingfield et al. 1990), and higher testosterone levels have been shown to reduce overall fitness in some avian species (Wingfield et al. 2001); 5) some birds habituate to recorded calls and greatly reduce their response rates over time (Harris and Haskell 2013), thus decreasing detection probabilities when the same recording is used over long time periods; 6) response to these surveys is strongest during the breeding season when birds are paired and defending territories, thus, callbacks may also be less effective at detecting non-territorial individuals in a population, as demonstrated for great horned owls (*Bubo virginianus*) in the southwestern Yukon in Canada, and eagle owls (*B. bubo*) in Spain (Rohner 1997, Martínez and Zuberogitia 2003); and 7) broadcast surveys are species-specific, and any incidental detections of other species during those surveys may not accurately represent the occupancy patterns of those additional species. For example, detection rates of barred owls during callback surveys for NSO are lower than detection rates during callback surveys specifically focused on barred owls. Thus, naïve estimates of barred owl occupancy rates can be underestimated when based on incidental detections associated with NSO-specific callback surveys (Wiens et al. 2014).

There are potentially more severe consequences to the use of callback surveys for NSOs when barred owls are present, because barred owls will approach and may react aggressively to the source of a NSO broadcast call (Herter and Hicks 2000, Piorecky and Prescott 2004, Wiens et al. 2011). Considering interference competition (Wiens et al. 2014) and aggression observed between the two species (Leskiw and Gutiérrez 1998, Courtney et al. 2004, Gutiérrez et al. 2007, Van Lanen et al. 2011, Lesmeister et al. 2018), eliciting NSO responses in areas where barred owls occur can increase the negative interactions between the two species. Additionally, NSO respond to callbacks less frequently if barred owls are present (Crozier et al. 2006), requiring more surveyor effort to determine occupancy status on historic survey sites. Given the many potential risks associated with callback surveys and continued population declines of NSOs, it is increasingly important to evaluate alternative survey methods.

Passive acoustic monitoring using autonomous recording units (ARUs) is a fast-growing area of wildlife research (Laiolo 2010, Shonfield and Bayne 2017), and has been especially useful in surveys of species that are rare, nocturnal, and difficult to detect by traditional means (Blumstein et al. 2011). Studies in remote areas with difficult access have benefited greatly from recent advances in ARU technology as large amounts of data can be collected with little human effort. Advances in ARU technology now allow for extended-duration sessions, which greatly decreases technician effort in the field while greatly increasing the quantity of data collected (Tegeler et al. 2012). Several types of automated detection software are now available that employ sophisticated recognition systems and several frameworks have been developed to analyze large acoustic data sets (Blumstein et al. 2011, Katz et al. 2016, Kahl et al. 2017, Chambert et al. 2018, Ruff et al. *In Review*). Additionally, Wood et al. (2019) demonstrated the ability to monitor small changes in spotted owl populations using passive acoustic monitoring.

There are many advantages to using ARUs compared to vocal lure methods: 1) surveys are non-invasive and do not require an elicited response from the target species, eliminating the risk listed above; 2) fieldwork takes place during daylight hours; this is an important factor when surveying nocturnal species because day work greatly improves crew safety; 3) biological training and expertise needed for field crews is less than what is needed for vocal response surveys, point counts and demographic studies; 4) experts can check and verify species identification of recordings back in the lab, decreasing issues of observer bias in the field; 5) increased temporal coverage when deploying ARUs for long durations; 6) records all sounds, providing a rich data set for multi-species monitoring purposes; and 7) resulting sound data provide a permanent record of all vocal species at each ARU location, dramatically broadening the scope for future analyses.

While there are many advantages for using ARUs as a monitoring tool, there are challenges associated with these methods: 1) there can be a high initial cost of investment in ARU hardware and development of automated call extraction techniques, but using ARUs over the long-term on large-scale projects is most cost-effective due to the decrease in the number of field crew requirements and training; 2) ARUs can accumulate a large volume of sound file data, so storing and in particular, processing the data may be costly and logistically challenging; and 3) models are still in development for using detection/non-detection data (as collected with ARUs) to estimate apparent survival and reproduction.

Study Objectives

Here we provide a progress report on passive bioacoustic research conducted during 2017–2019 using a combination of ARU surveys and deep learning methods to effectively detect and monitor NSOs and barred owls across a large geographic area. We present a sampling methodology informed by recent findings on calling activity patterns of both species. Our primary objectives for this research were to 1) quantify seasonal and diel calling activity of NSOs and barred owls; 2) quantify factors that most affect NSO detection probabilities, including factors that may affect calling behavior and ability of ARUs to record calling when it occurs; 3) determine the necessary number and placement of ARUs within a 500-ha hexagon to achieve a seasonal detection probability >0.85 for NSO using 1-week intervals as sampling occasions; 4) quantify the presence and vocal activity of NSOs and barred owls in a random sample of field sites within the Olympic Peninsula and Oregon Coast Range study areas; and 5) develop an effective deep-learning model that could efficiently automate the detection of NSO

and barred owl vocal activity in the large volumes of acoustic data generated by broad-scale ARU deployments, with only a nominal level of human effort required to validate model output.

4. Study Area

We collected data for objectives one–three during 2017 on three long-term NSO demographic study areas (KLA = Klamath, COA = Coast Range, and OLY = Olympic Peninsula), and during 2018 we collected data for objective four in COA and OLY. Lands were under federal ownership administered by US Forest Service, US Bureau of Land Management, and National Park Service, which make up a portion of the lands surveyed as part of the NSO Effectiveness Monitoring Program under the Northwest Forest Plan (Fig. 1). These three study areas have been included in the numerous meta-analyses conducted to understand the status and trends of NSO across the species range (Anthony et al. 2006, Forsman et al. 2011, Dugger et al. 2016). In 2019, we are collecting data from broad-scale ARU surveys in all three study areas.

5. Methods

Sampling design

We collected data using Wildlife Acoustics Song Meter SM4s (www.wildlifeacoustics.com), which are portable, weatherproof, and easily programmable, with two built-in high-quality microphones, large memory capacity, 350-400 hour battery life, and the capacity to record sound between 20 Hz and 48 kHz at decibel levels of -33.5 dB to 122 dB (Wildlife Acoustics Inc. 2017). ARUs record sound with equivalent sensitivity to human hearing, and their effective listening radius may be affected by external factors such as terrain, vegetation, and weather events such as wind and rain. SM4s and earlier Song Meter models have been widely used to monitor a range of amphibian, avian and mammalian species (e.g., Zwart et al. 2014, Sidie-Slettedahl et al. 2015, Courtois et al. 2016, Kalan et al. 2016, Ross et al. 2018).

We created a uniform layer of 5-km^2 hexagons that covered the entire range of the NSO. This hexagon size approximately the size of a NSO territory core area centered on a primary activity center or nest tree (Glenn et al. 2004, Schilling et al. 2013) and is the mean home range size reported for barred owls in the Pacific Northwest (Hamer et al. 2007, Singleton et al. 2010). Thus, hexagons of this size reflect ecologically relevant space use by both spotted and barred owls during the breeding season. Spotted owl and barred owl territories are likely to overlap more than one hexagon; therefore, we did not survey adjacent hexagons to minimize detections of the same owl in multiple hexagons. Surveying non-adjacent 5-km^2 hexagons provides a buffer between territories and reduces detecting the same individual in multiple hexagons. To be considered as a sample, the hexagon needed to contain ≥50% forest capable lands and ≥25% federal ownership.

2017 data collection

During the 2017 NSO breeding season (March–August) we deployed ARUs at 150 recording stations in 30 sample hexagons (5 ARUs/hexagon) that had NSO nesting and activity histories during 2016. We surveyed 10 hexagons in each study area. During 2017 ARUs recorded continuously from 1 h before sunset to 2 h after sunrise each night, producing 11 – 15 h of recordings per night over the season. ARUs recorded in stereo at a rate of 32 kHz. ARUs were deployed for ~4 months at each location. After deployment, crews visited ARUs once every 4

weeks for battery and SD card replacement, and retrieved ARUs during late June–mid-July. We mounted ARUs to small trees with diameter ~15-20cm to allow microphones to extend past the trunk for unobstructed recording ability. We positioned ARUs in mid-to-upper slope positions and at least 50 meters from roads, trails and streams to reduce vandalism and excessive noise. Each ARU holds one 512 GB SD memory card and 4 D batteries, which power the ARUs for approximately 350 hours of recording.

2018 data collection

During the 2018 NSO breeding season we deployed ARUs at 1,040 recording stations in 208 sample hexagons (5 units/hexagon) in COA and OLY (Fig. 2). We intended to deploy ARUs in 120 hexagons in each study area; however, due to poor road access early in the season, only 88 hexagons were ultimately sampled in OLY. During 2018 ARUs recorded from 1 h before sunset to 3 h after sunset and from 2 h before sunrise to 2 h after sunrise, producing 8 h of recordings per night. ARUs recorded in stereo at a rate of 32 kHz. Based on analysis of 2017 data, we found that with 6 weeks of sampling per site we obtained robust detection probabilities for NSO of ~0.98, therefore, hexagons were surveyed for 6 weeks during 2018. ARUs were deployed within each hexagon as described above (2017).

2019 data collection

In 2019 we are deploying ARUs in 120 hexagons in COA, 120 in OLY, and 73 in KLA, for a planned total of 1,252 recording stations across all 3 study areas. Based on findings from 2017 and 2018 we found that surveying with 4 stations/hexagons resulted in seasonal detection probabilities of ~0.98, thus we used 4 ARUs per hexagon in 2019. ARU deployments were for 6 weeks, as in 2018. This deployment schedule is sufficient to obtain robust detection probability at the hexagon scale. During 2019 ARUs are recording on the same crepuscular schedule but additionally will record for the first 10 m of every hour throughout the day and night, allowing for additional detections of diurnally-active species. In 2019 ARUs will record in mono to produce a smaller volume of data. ARUs were deployed as in previous years.

Analysis of 2017 data

We extracted owl vocalizations from the 2017 data using the basic clustering feature in Kaleidoscope Pro software, which detects sounds matching user-defined criteria and clusters them according to similarity using a hidden Markov model (Wildlife Acoustics 2017). We then manually reviewed these sounds to find territorial calls of NSOs and barred owls. Overall we detected 20,312 NSO vocalizations and 67,734 barred owl vocalizations in the 2017 data. Using these detections, we built encounter histories and performed several occupancy analyses to examine biotic and abiotic factors affecting detection probability of both species.

We used single-season, single-species and single-season co-occurrence occupancy models (MacKenzie et al. 2018) in the R package RPresence (R Core Team 2019) to estimate NSO occupancy or use rates (ψ) and detection probabilities (p) for NSOs and barred owls. For objectives one–three we focused on estimating p for NSO and barred owls (p_{NSO} and p_{BO}) at two spatial scales (ARU- and hexagon-level). We used an information theoretic approach to evaluate model support for *a priori* models that included covariates hypothesized to affect p_{NSO} and p_{BO} (Table 1), and ranked models using Akaike's information criterion corrected for small sample size (AIC_c) (Burnham and Anderson 2002). Using the best model in the single-species NSO

model set, we calculated seasonal power of detection to determine the minimum number of ARUs and survey weeks required per hexagon to achieve a seasonal NSO detection probability of ≥0.95.

Automation of NSO and barred owl detections

We used the vocalizations of NSOs and barred owls from 2017 data as part of a training data set to train a convolutional neural network (CNN) model (Ruff et al. *In Review*), which we are currently using to detect stereotypic calls of 12 avian and two mammalian species, including NSOs and barred owls, in the 2018 data. In addition to NSO and barred owls, we trained the CNN to recognize typical vocalizations of northern saw-whet owl *Aegolius acadicus*, great horned owl, common raven *Corvus corax*, Steller's jay *Cyanocitta stelleri*, northern pygmy-owl *Glaucidium gnoma*, pileated woodpecker *Hylatomus pileatus*, western screech-owl *Megascops kennicottii*, mountain quail *Oreortyx pictus*, band-tailed pigeon *Patagioenas fasciata*, red-breasted sapsucker *Sphyrapicus ruber*, Douglas' squirrel *Tamasciurus douglasii*, and Townsend's chipmunk *Tamias townsendii*. The inclusion of additional target classes beyond NSOs and barred owls, particularly those whose vocalizations are similar to one another, improves the model's discriminative ability and improves accuracy. Although some of these target species are of marginal interest, the extraction of these detections demonstrates the robustness of passive bioacoustics for multi-species monitoring and community-level analyses.

The CNN is an image classification model, therefore we split all sound files (.wav) into 12 second segments and then converted those to spectrograms, which are image representations of sound (Fig. 3). We implemented the CNN model in Python using Keras (Chollet 2015), an API to Google's TensorFlow software library (Abadi et al. 2015). Details on development of the CNN can be found in Ruff et al. (2018), Ruff et al. (*In Review*), and Appendix A. We trained the CNN for 100 epochs on a set of 173,964 images, of which 80% were used for training and 20% were set aside as a validation set. The output from the CNN are probabilities (from 0 to 1.0) that each 12-second segment of sound is of each of the 15 identified classes (i.e., 14 species and background noise). All output from the CNN was validated by trained human reviewers in order to generate species encounter histories to be used in occupancy models and other analyses. For NSO we validate all clips that scored higher than a 0.25 probability regardless of what the highest-scoring class was, and all clips that scored higher than 0.95 for other target species. We also validate a random sample of 0.5% of all other clips. These rules are intended to maximize detection power for NSOs, allow us to quickly produce encounter histories for all target species at the hexagon level, and minimize time spent on false positives. Validating a random sample of clips that did not otherwise merit consideration allows us to estimate the number of false negatives, which were the number of real vocalizations that would otherwise be missed under this validation scheme.

6. Results

2017 findings

During 2017, we collected 67 Terabytes (TB) of sound data over the survey period, with ~25 TB each from COA and KLA, and ~17 TB from OLY, totaling over 150,000 hours (equals

about 17 years) of sound. We identified NSO vocalizations in 26 out of 30 hexagons, and barred owls in 28 out of 30 hexagons. Clustering also revealed the calls of northern pygmy-owls, western screech-owls, great horned owls, and northern saw-whet owls. Using all positively identifiable vocalizations from barred owls and NSOs, we observed differences in calling activity patterns: NSOs were most actively calling during the crepuscular periods, around sunset and sunrise, whereas barred owls were most active during the nocturnal period (Fig. 4). We also observed increased vocal activity for NSOs and barred owls near the full moon phase (Fig. 5).

Results from detection probability models showed that background noise from rain, streams, and wind strongly influenced the ability of ARUs to detect both NSO and barred owl vocalizations (Fig. 6). Additionally, detection probability varied between study areas, with the highest probability of detection in KLA for NSOs and COA for barred owls, and lowest in OLY for both species. Based on information from concurrent demographic surveys, we found that known pairs of NSOs had much higher detection probabilities (approaching 1) than known single NSOs on the landscape. In co-occurrence occupancy models of barred owls and NSOs, NSOs had the highest detection probabilities at ARU stations where NSOs were present but barred owls were never detected over the survey season. However, this was rare, only occurring at 7 of 150 ARU stations. At stations where both species were present, NSO detection probability was significantly higher in weeks when barred owls also vocalized, suggesting that at these sites, NSOs were actively defending their territories against barred owl incursions.

While weekly detection probability at a single ARU station varied, summarizing detections to the hexagon scale resulted in robust detection probabilities in all study areas. Using a reduced 4-ARU design and full-night recording for sampling a hexagon, seasonal probability of detection exceeded 0.95 for NSOs within 2 weeks of surveys at KLA and 3 weeks of surveys at COA and OLY. Barred owl seasonal probability of detection using 4-ARU design exceeded 0.95 for all study areas within 1-2 weeks.

CNN model performance

During training the CNN reached a validation accuracy of 99.69% and a validation loss of 0.0139. Using a "naïve" classification scheme in which each clip was classified as the target class with the highest predicted score, precision was 74.0% for NSO four-note calls, 81.3% for barred owl eight-note calls, and 79.2% for barred owl inspection calls. Under the same scheme, recall was 95.9% for NSO four-note calls, 91.9% for barred owl eight-note calls, and 95.8% for barred owl inspection calls. We can achieve greater precision by considering only apparent detections for which the top class score exceeds some threshold. For example, if we consider only clips for which the top predicted class score is ≥ 0.9, precision would be 92.9% for NSO, 91.6% for barred owl eight-note calls, and 96.6% for barred owl inspection calls. Applying this threshold lowers recall, which would be 82.8% for NSO, 81.6% for barred owl eight-note calls, and 84.3% for barred owl inspection calls. There is some tradeoff between precision and recall; models with high specificity tend to be less sensitive and vice versa.

2018 preliminary results

During 2018 we collected approximately 150 terabytes (TB) of acoustic data across two study areas, with ~86 TB from COA and ~64 TB from OLY, totaling approximately 350,000 hours (equally about 40 years) of sound. We began using the CNN to process this data in late

May 2019 and as of June 20, 2019 we have since processed data from 42 of the 208 hexagons sampled in 2018, totaling approximately 71,000 hours of recordings. Technicians are currently validating output from the CNN, prioritizing clips tagged as NSO and barred owl calls. We have identified short bouts of NSO vocalizations at several sites. We have confirmed the presence of barred owls at all sites validated thus far.

2019 field season accomplishments

An eight-person crew in the OLY is beginning retrieval after deploying ARUs in 72 hexagons (288 ARUs, 4 per hexagon) and will re-deploy ARUs in an additional 48 hexagons. A four-person crew in COA has deployed ARUs and co-located remote cameras in 50 hexagons (200 ARUs) and continues to deploy with a goal of 120 hexagons over the season, and a two-person crew in KLA has completed their first round of deployments of 36 hexagons and will begin retrieving and re-deploying ARUs in the second week of June, 2019.

7. Discussion

Results contained in this report are from ongoing research and as such are preliminary and should not be used to inform land-management decision making. Our approach to passive acoustic monitoring has enabled data collection at an unprecedented scale, allowing us to sample across a large portion of the NSO's range with fewer field personnel and improved crew safety. Our results support findings from other researchers (e.g., Wood et al. 2019) that found passive acoustic monitoring to be a highly effect method to monitor the status and trends in spotted owl populations. Our development of the CNN to automate the detection of NSOs and barred owls has taken approximately 16 months to complete, but is now fully operational with rates of precision and recall for species identification that are higher than is reported elsewhere in the scientific literature, for any species. For example, see Knight et al. (2017) and Shonfield et al. (2018). The CNN has already been highly effective at automating the detection of target species, especially NSOs and barred owls. Although further development will be necessary, we are confident that the CNN will reliably—and rapidly—detect NSOs and barred owls, as well as other target species present at our field sites. Further, the sensitivity of our methods will facilitate the quantification of number of vocalizations with sufficient power to allow for accurate characterizations of activity level. This will facilitate inference about pair or nesting status of NSOs, greatly increasing our ability to understand population dynamics using passive acoustic monitoring.

8. Acknowledgments

Funding was provided by USDI Bureau of Land Management and USDA Forest Service. We are thankful for the hard work and dedication of David Culp and others who have provided field assistance, including Clara Cardillo, Maggie Corr, Taylor Garrido, Eric Guzman, Angela Ingrassia, DeAnne Jacobsma, Erin Johnston, Robert Justice, Debaran Kelso, Katie McLaughlin, and Philip Papajcik. We are deeply indebted to Chris Sullivan and Bharath Padmaraju for assistance with developing and training the convolutional neural network model. We thank Katie Dugger for many brainstorm sessions on using bioacoustics to monitor spotted owls, and for serving as L. Duchac's co-advisor. Various levels of logistical and permitting assistance was provided by Chris Foster, Scott Gremel, Patti Happe, Rob Horn, Krista Lewicki, Chris

McCafferty, Donna Owen, Susan Piper, Shane Pruett, Janice Reid, Errin Trujillo, and Kari Williamson; thank you.

9. Literature Cited

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. T. Steiner, P., V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, and Google Brain. 2015. TensorFlow: A System for Large-Scale Machine Learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16).

Anthony, R. G., E. D. Forsman, A. B. Franklin, D. R. Anderson, K. P. Burnham, G. C. White, C. J. Schwarz, J. D. Nichols, J. E. Hines, G. S. Olson, S. H. Ackers, L. S. Andrews, B. L. Biswell, P. C. Carlson, L. V. Diller, K. M. Dugger, K. E. Fehring, T. L. Fleming, R. P. Gerhardt, S. A. Gremel, R. J. Gutierrez, P. J. Happe, D. R. Herter, J. M. Higley, R. B. Horn, L. L. Irwin, P. J. Loschl, J. A. Reid, and S. G. Sovern. 2006. Status and trends in demography of northern spotted owls, 1985-2003. Wildlife Monographs 163:1-48.

Blumstein, D. T., D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, S. F. Hanser, B. McCowan, A. M. Ali, and A. N. G. Kirschel. 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. Journal of Applied Ecology 48:758-767.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd edition. Springer, New York, NY.

Chambert, T., J. H. Waddle, D. A. W. Miller, S. C. Walls, J. D. Nichols, and N. Yoccoz. 2018. A new framework for analysing automated acoustic species detection data: Occupancy estimation and optimization of recordings post-processing. Methods in Ecology and Evolution 9:560-570.

Chollet, F. o. 2015. Keras. https://github.com/fchollet/keras.

Conway, C. J., V. Garcia, M. D. Smith, and K. Hughes. 2008. Factors affecting detection of burrowing owl nests during standardized surveys. Journal of Wildlife Management 72:688-696.

Conway, C. J., and J. P. Gibbs. 2005. Effectiveness of call-broadcast surveys for monitoring marsh birds. The Auk 122:26-35.

Courtney, S. P., J. A. Blakesley, R. E. Bigley, M. L. Cody, J. P. Dumbacher, R. C. Fleishcher, A. B. Franklin, J. F. Franklin, R. J. Gutiérrez, J. M. Marzluff, and L. Sztukowski. 2004. Scientific evaluation of the status of the northern spotted owl. Sustainable Ecosystems Institstute.

Courtois, E. A., E. Michel, Q. Martinez, K. Pineau, M. Dewynter, G. F. Ficetola, and A. Fouquet. 2016. Taking the lead on climate change: modelling and monitoring the fate of an Amazonian frog. Oryx 50:450-459.

Crozier, M. L., M. E. Seamans, R. J. Gutierrez, P. J. Loschl, R. B. Horn, S. G. Sovern, and E. D. Forsman. 2006. Does the presence of barred owls suppress the calling behavior of spotted owls? Condor 108:760-769.

Dugger, K. M., E. D. Forsman, A. B. Franklin, R. J. Davis, G. C. White, C. J. Schwarz, K. P. Burnham, J. D. Nichols, J. E. Hines, C. B. Yackulic, P. F. Doherty Jr., L. L. Bailey, D. A. Clark, S. H. Ackers, L. S. Andrews, B. Augustine, B. L. Biswell, J. A. Blakesley, P. C. Carlson, M. J. Clement, L. V. Diller, E. M. Glenn, A. Green, S. A. Gremel, D. R. Herter,

J. M. Higley, J. Hobson, R. B. Horn, K. P. Huyvaert, C. McCafferty, T. L. McDonald, K. McDonnell, G. S. Olson, J. A. Reid, J. Rockweit, V. Ruiz, J. Saenz, and S. G. Sovern. 2016. The effects of habitat, climate and Barred Owls on the long-term population demographics of Northern Spotted Owls. Condor 118:57-116.

Forsman, E. D., R. G. Anthony, K. M. Dugger, E. M. Glenn, A. B. Franklin, G. C. White, C. J. Schwartz, K. P. Burnham, D. R. Anderson, J. D. Nichols, J. E. Hines, J. B. Lint, R. J. Davis, S. H. Ackers, L. S. Andrews, B. L. Biswell, P. C. Carlson, L. V. Diller, S. A. Gremel, D. R. Herter, J. M. Higley, R. B. Horn, J. A. Reid, J. Rockweit, J. P. Schaberl, T. J. Snetsinger, and S. G. Sovern. 2011. Population Demography of Northern Spotted Owls. Studies in Avian Biology 40:1-106.

Glenn, E. M., M. C. Hansen, and R. G. Anthony. 2004. Spotted owl home-range and habitat use in young forests of western Oregon. Journal of Wildlife Management 68:33-50.

Gremel, S. A. 2019. Spotted Owl Monitoring in Olympic National Park: 1992-2018. National Park Service. Report Report NPS/OLYM/NRR—2017/XXX.

Gutiérrez, R. J., M. Cody, S. Courtney, and A. B. Franklin. 2007. The invasion of barred owls and its potential effect on the spotted owl: A conservation conundrum. Biological Invasions 9:181-196.

Hamer, T. E., E. D. Forsman, and E. M. Glenn. 2007. Home range attributes and habitat selection of Barred Owls and Spotted Owls in an area of sympatry. Condor 109:750-768.

Harris, J. B., and D. G. Haskell. 2013. Simulated birdwatchers' playback affects the behavior of two tropical birds. Plos One 8:e77902.

Haug, E. A., and A. B. Didiuk. 1993. Use of recorded calls to detect burrowing owls. Journal of Field Ornithology 64:188-194.

Herter, D. R., and L. L. Hicks. 2000. Barred owl and spotted owl populations and habitat in the central Cascade Range of Washington. Journal of Raptor Research 34:279-286.

Kahl, S., T. Wilhelm-Stein, H. Hussein, H. Klinck, D. Kowerko, M. Ritter, and M. Eibl. 2017. Large-scale bird sound classification using convolutional neural networks. *in* BirdCLEF 2017.

Kalan, A. K., A. K. Piel, R. Mundry, R. M. Wittig, C. Boesch, and H. S. Kuhl. 2016. Passive acoustic monitoring reveals group ranging and territory use: a case study of wild chimpanzees (Pan troglodytes). Frontiers in Zoology 13:34.

Katz, J., S. D. Hafner, and T. Donovan. 2016. Tools for automated acoustic monitoring within the R package monitoR. Bioacoustics 25:197-210.

Knight, E. C., K. C. Hannah, G. Foley, C. Scott, R. M. Brigham, and E. Bayne. 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conservation and Ecology 12:Article 14.

Laiolo, P. 2010. The emerging significance of bioacoustics in animal species conservation. Biological Conservation 143:1635-1645.

Langham, G. M., T. A. Contreras, and K. E. Sieving. 2006. Why pishing works: Titmouse (Paridae) scolds elicit a generalized response in bird communities. Ecoscience 13:485-496.

Leskiw, T., and R. J. Gutiérrez. 1998. Possible predation of a spotted owl by a barred owl. Western Birds 29:225-226.

Lesmeister, D. B., R. J. Davis, P. H. Singleton, and J. D. Wiens. 2018. Northern spotted owl habitat and populations: status and threats *in* T. Spies, P. Stine, R. Gravenmier, J. Long,

andM. Reilly, editors. Synthesis of Science to Inform Land Management within the Northwest Forest Plan Area. PNW-GTR-966. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

Lesmeister, D. B., M. S. Pruett, D. Kelso, and K. Williamson. 2019. Demographic Characteristics of Northern Spotted Owls (*Strix occidentalis caurina*) in the Olympic National Forest, Washington, 1987–2018. USDA Forest Service, Pacific Northwest Research Station.

Lint, J., B. Noon, R. Anthony, E. Forsman, M. Raphael, M. Collopy, and E. Starkey. 1999. Northern Spotted Owl Effectiveness Monitoring Plan for the Northwest Forest Plan. USDA Forest Service, Pacific Northwest Research Station. Report PNW-GTR-440.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2018. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. 2nd Edition. . 2nd edition. Academic Press, Cambridge, MA.

Martínez, J. A., and I. Zuberogitia. 2003. Factors afecting the vocal behaviour of eagle owls *Bubo bubo*: Effects of season, density and territory quality. Ardeola 50:255-258.

Mennill, D. J., L. M. Ratcliffe, and P. T. Boag. 2002. Female eavesdropping on male song contests in songbirds. Science 296:873.

Moulton, C. E., R. S. Brady, and J. R. Belthoff. 2004. Territory defense of nesting Burrowing Owls: responses to simulated conspecific intrusion. Journal of Field Ornithology 75:288-295.

Piorecky, M. D., and D. R. C. Prescott. 2004. Distribution, abundance and habitat selection of northern pygmy and barred owls along the eastern slopes of the Alberta Rocky Mountains. Alberta Species at Risk Report No. 91. Alberta Sustainable Resource Development, Fish and Wildlife Division, Edmonton, AB.

R Core Team. 2019. Version 3.3.1. R Foundation for Statistical Computing, Vienna, Austria.

Rohner, C. 1997. Non-territorial 'floaters' in great horned owls: space use during a cyclic peak of snowshoe hares. Animal Behaviour 53:901-912.

Ross, S. R. P. J., N. R. Friedman, K. L. Dudley, M. Yoshimura, T. Yoshida, and E. P. Economo. 2018. Listening to ecosystems: data-rich acoustic monitoring through landscape-scale sensor networks. Ecological Research 33:135-147.

Ruff, Z., B. K. Padmaraju, C. M. Sullivan, and D. B. Lesmeister. 2018. Spotted Owls and OpenPower. *in* Linux on Power Developer Portal. IBM Developer, https://developer.ibm.com/linuxonpower/2018/11/21/spotted-owls-and-openpower/.

Ruff, Z. J., D. B. Lesmeister, L. S. Duchac, B. K. Padmaraju, and C. M. Sullivan. *In Review*. Automatic identification of avian vocalizations with deep convolutional neural networks. Remote Sensing in Ecology and Conservation.

Schilling, J. W., K. M. Dugger, and R. G. Anthony. 2013. Survival and home-range size of northern spotted owls in southwestern Oregon. Journal of Raptor Research 47:1-14.

Shonfield, J., and E. M. Bayne. 2017. Autonomous recording units in avian ecological research: current use and future applications. Avian Conservation and Ecology 12:14.

Shonfield, J., S. Heemskerk, and E. M. Bayne. 2018. Utility of automated species recognition for acoustic monitoring of owls. Journal of Raptor Research 52:42-55.

Sidie-Slettedahl, A. M., K. C. Jensen, R. R. Johnson, T. W. Arnold, J. E. Austin, and J. D. Stafford. 2015. Evaluation of autonomous recording units for detecting 3 species of secretive marsh birds. Wildlife Society Bulletin 39:626-634.

Singleton, P. H., J. F. Lehmkuhl, W. L. Gaines, and S. A. Graham. 2010. Barred owl space use and habitat selection in the eastern Cascades, Washington. Journal of Wildlife Management 74:285-294.

Tegeler, A. K., M. L. Morrison, and J. M. Szewczak. 2012. Using extended-duration audio recordings to survey avian species. Wildlife Society Bulletin 36:21-29.

Van Lanen, N. J., A. B. Franklin, K. P. Huyvaert, R. F. Reiser, and P. C. Carlson. 2011. Who hits and hoots at whom? Potential for interference competition between barred and northern spotted owls. Biological Conservation 144:2194-2201.

Wiens, J. D., R. G. Anthony, and E. D. Forsman. 2011. Barred owl occupancy surveys within the range of the northern spotted owl. The Journal of Wildlife Management 75:531-538.

Wiens, J. D., R. G. Anthony, and E. D. Forsman. 2014. Competitive interactions and resource partitioning between northern spotted owls and barred owls in Western Oregon. Wildlife Monographs 185:1-50.

Wingfield, J. C., R. E. Hegner, A. M. Dufty, and G. F. Ball. 1990. The "challenge hypothesis": Theoretical implications for patterns of testosteronesecretion, mating systems, and breeding strategies. The American Naturalist 136:829-846.

Wingfield, J. C., S. E. Lynn, and K. K. Soma. 2001. Avoiding the 'costs' testosterone: Ecological bases of hormone-behavior interactions. Brain, Behavior and Evolution 57:239-251.

Wood, C. M., V. D. Popescu, H. Klinck, J. J. Keane, R. J. Gutiérrez, S. C. Sawyer, and M. Z. Peery. 2019. Detecting small changes in populations at landscape scales: a bioacoustic site-occupancy framework. Ecological Indicators 98:492-507.

Worthington-Hill, J., and G. Conway. 2017. Tawny Owl Strix aluco response to call-broadcasting and implications for survey design. Bird Study 64:205-210.

Zwart, M. C., A. Baker, P. J. McGowan, and M. J. Whittingham. 2014. The use of automated bioacoustic recorders to replace human wildlife surveys: an example using nightjars. Plos One 9:e102770.

10. Tables

Table 1. Site- and survey-specific covariates for autonomous recording unit locations and hexagons, to model detection probabilities in three northern spotted owl demographic study areas for the 2017 breeding season.

Variable	Description
Precipitation	Daily precipitation in centimeters derived from PRISM climate data (PRISM Climate Group 2018, prismclimate.org) averaged daily per ARU location. PRISM precipitation data are reported in a rectangular grid with units of 0.63km^2 or 63 ha. Approximately 14 PRISM grid units fall fully or partly within each 500-ha hexagon.
Terrain ruggedness	Defined as standard deviation of elevation in the area within each hexagon. Derived from GIS data.
Audible road	Binary, 1 if road noise is audible from ARU location—recorded in field upon deployment.
Topographic position	Topographic position within 450-m radius; continuous variable of relative position on slope, with the mid-slope as zero, upper slope values positive, and lower-slope values negative. Derived from GIS data.
Study Area	Categorical variable of 3 demographic study areas.
Background Noise	Measure of average weekly background noise level in dB from each ARU location between 220Hz and 1000Hz. Calculated by Kaleidoscope Pro software.
Distance to stream	Distance in meters from ARU location to nearest stream—derived from GIS.
Audible stream	Binary, 1 if stream noise is audible from during visit to ARU location—recorded in the field upon deployment.
Temperature	Hexagon-level temperature data collected every 2 hours by HOBO data loggers deployed on the ARU nearest to center of each hexagon, averaged weekly.
Week	Numbered week (1-18) of the survey season.
Distance to known NSO location	Distance from ARU location to nearest 2017 NSO nest location (if nesting) or estimated activity center (if not nesting)—from demographic study survey data.

NSO Pair Status	Pair status from demographic surveys of northern spotted owl territories—assigned from territory that overlapped each survey station: 0=no owls detected, 1=single owl detected, 2=non-nesting pair, 3=nesting pair.
Barred owl	Encounter history of barred owl detections at ARU and hexagon scale—derived from ARU data.

11. Figures

Figure 1. The Coast Range, Olympic, and Klamath northern spotted owl demographic study areas, in which autonomous recording units were deployed in 2017 and 2019. During 2018, units were deployed in Coast Range and Olympic study areas.

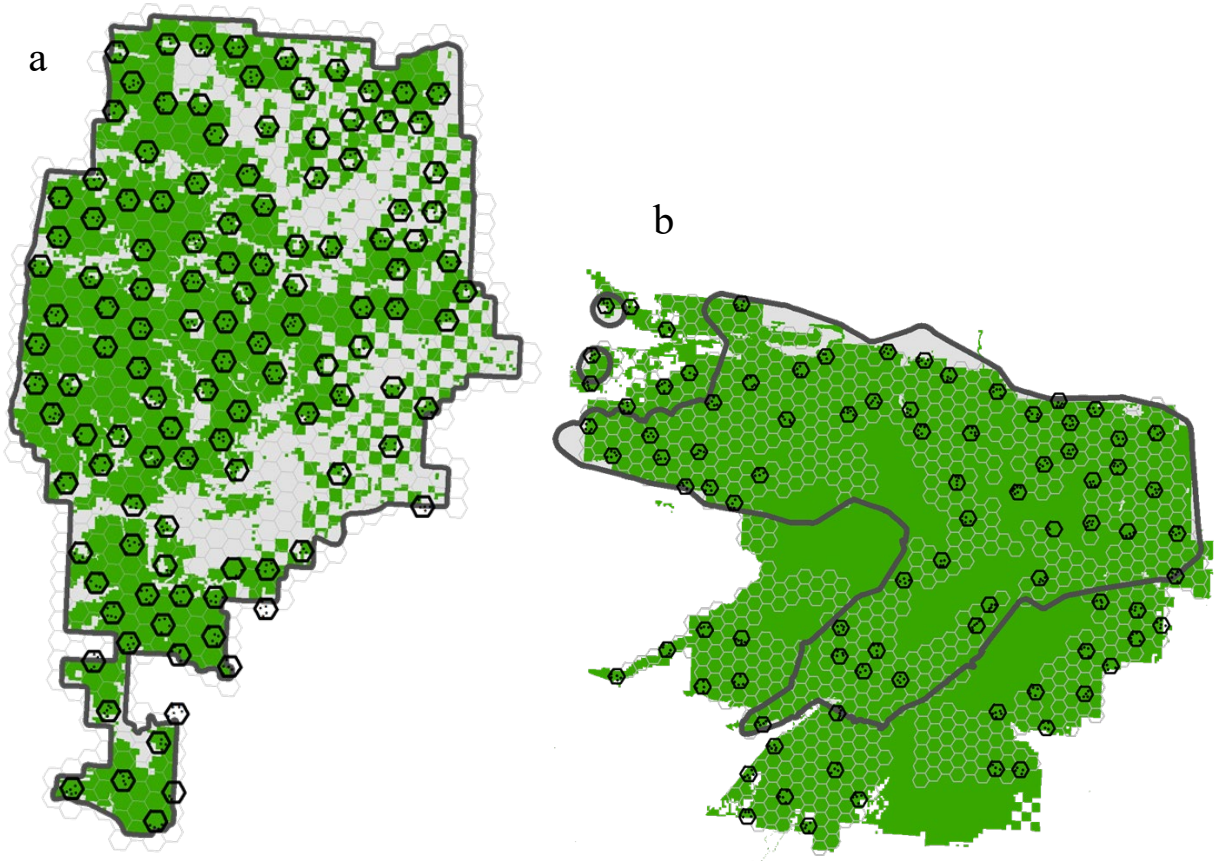


Figure 2. Maps of 5-km^2 hexagons (heavy black line) and ARU stations in a) Coast Range (n=120) and b) Olympic Peninsula (n=88) study areas surveyed during 2018. Demographic study area boundaries are shown in heavy gray line, with federal lands shown in green.

a

b

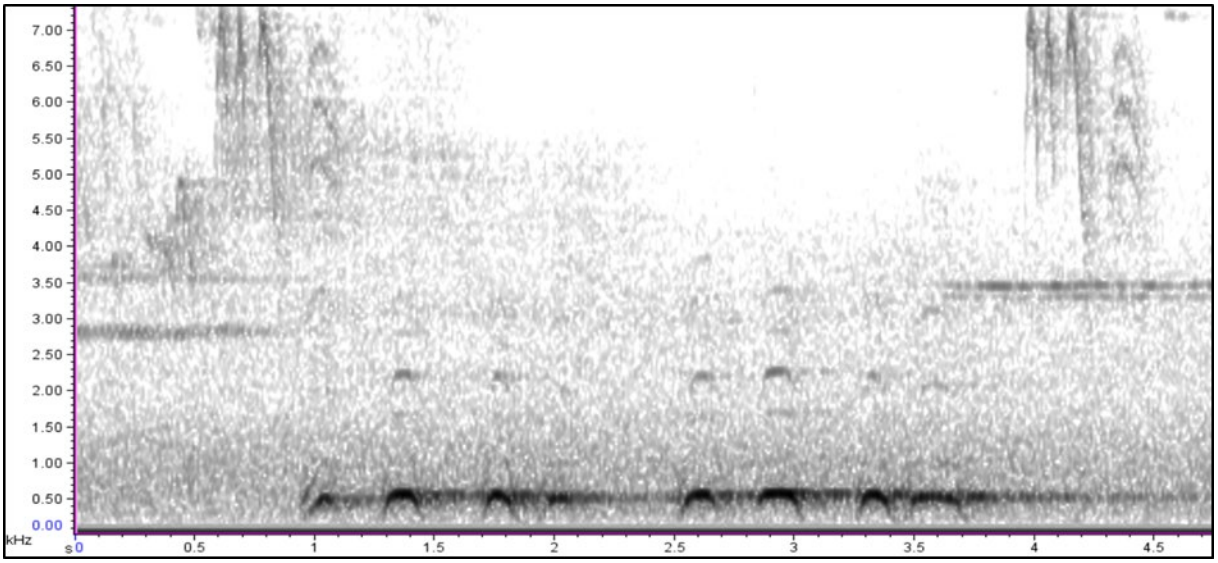


Figure 3. Spectrograms of a) northern spotted owl 4-note call and b) barred owl 8-note call.

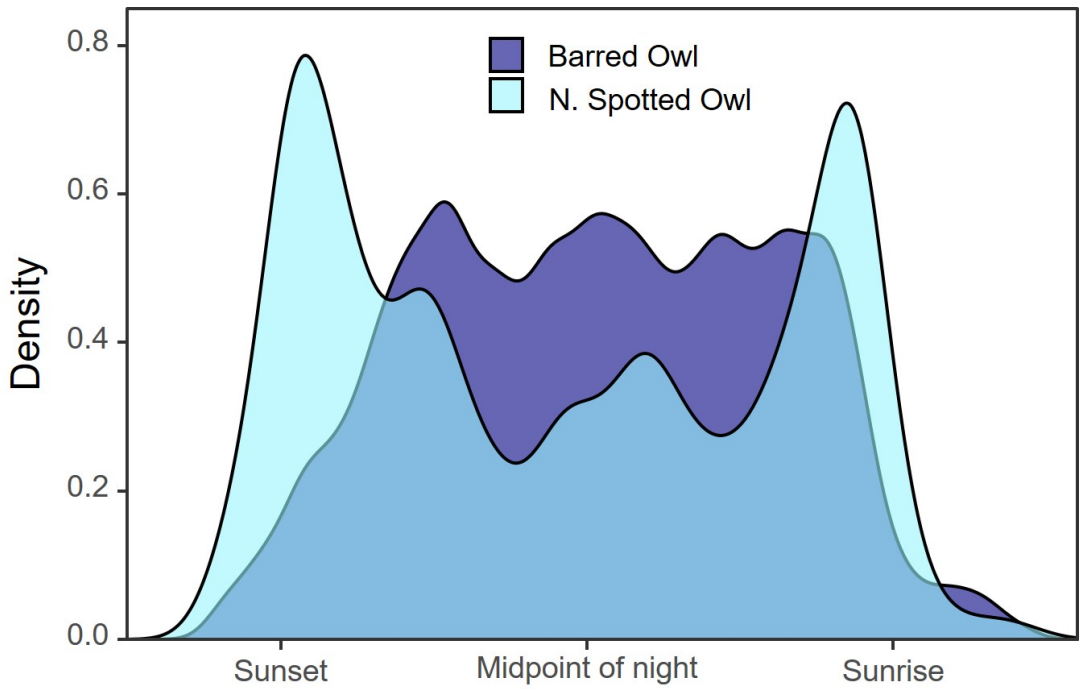


Figure 4. Density plot of daily calling activity based on ca. 20,000 northern spotted owl calls and 67,000 barred owl calls detected in data collected from study areas in Oregon and Washington in 2017. Time stamps of calls are normalized to reflect changes in night length over the course of the season.

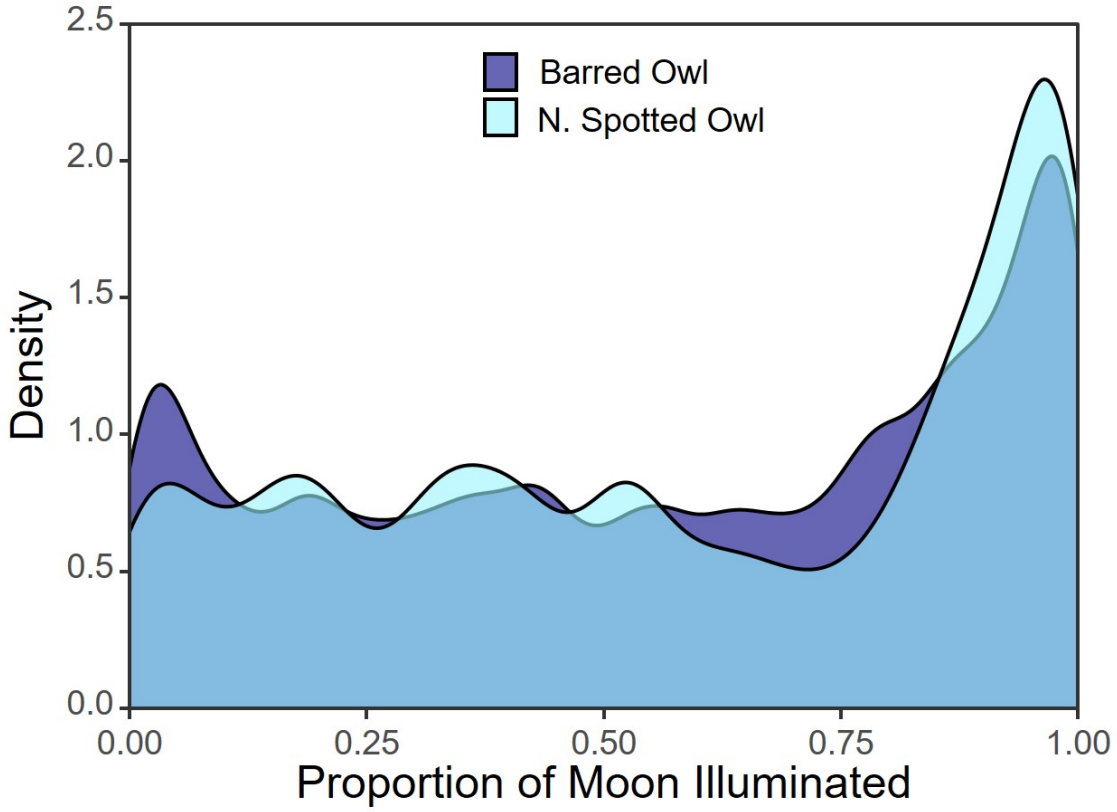
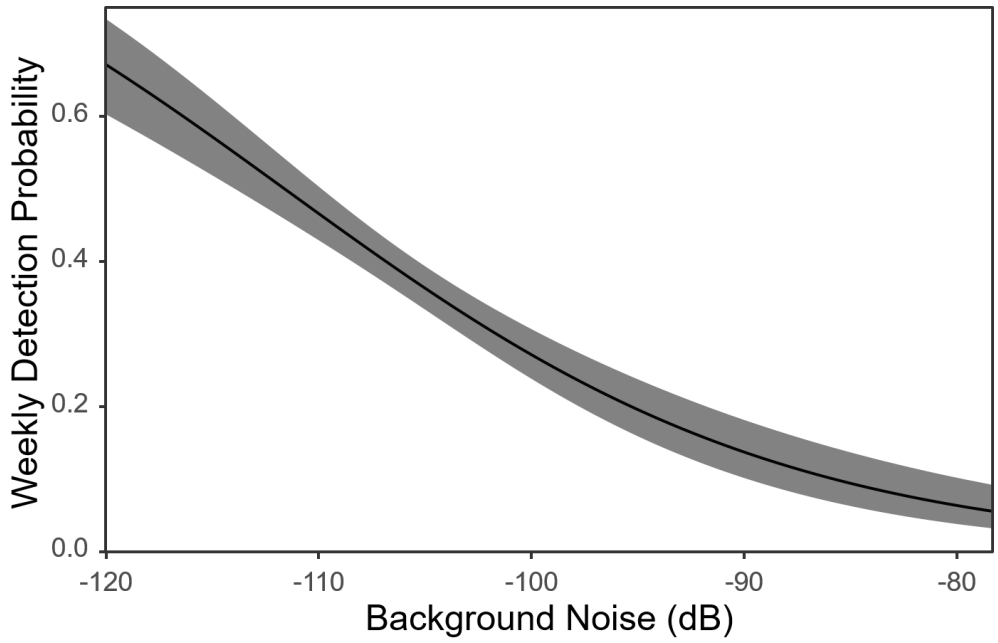


Figure 5. Density plot of calling activity based on ca. 20,000 northern spotted owl calls and 67,000 barred owl calls collected from three study areas in Oregon and Washington in 2017 in relation to moon phase, where full moon = 1.00.

a



b

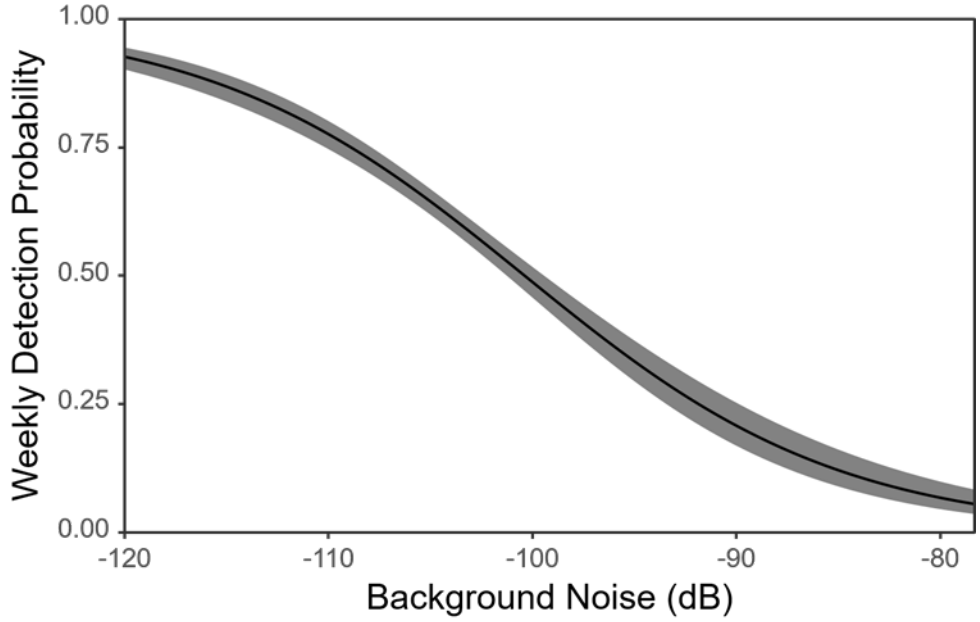


Figure 6. Detection probability relative to background noise for a) northern spotted owls and b) barred owls. Background noise is given as decibels below full scale (dBFS). Background noise was measured at frequencies from 250 – 1000 Hz.

12. Appendix A. Methods on the development of the Convolutional Neural Network (CNN) to automate the detection of northern spotted owls and barred owls.

We implemented the CNN model in Python using Keras (Chollet 2015), an API to Google's TensorFlow software library (Abadi et al. 2015). The CNN contains four convolutional layers and two fully connected layers. The first convolutional layer contains 32 5x5 filters and accepts grayscale spectrogram image data as input. The second convolutional layer contains 32 3x3 filters. The third and fourth convolutional layers each contain 64 3x3 filters. Each convolutional layer uses Rectified Linear Units (ReLU) activation and is followed by 2x2 max pooling and 20% dropout. The first fully connected layer contains 256 units and uses ReLU activation and L^2 regularization (squared Euclidean norm of the weight matrix of the hidden layer, or the sum of all such squared norms, in the case of multiple hidden layers, and including the output layer) and is followed by 50% dropout. The second fully-connected layer contains 17 units with sigmoid activation. The activation of this final layer comprises the model output and is interpretable as class scores for each of our 14 target classes.

We trained the CNN for 100 epochs on a set of 173,964 images, of which 80% were used for training and 20% were set aside as a validation set. We trained on batches of 128 images using the Adam optimization algorithm with an initial learning rate of 0.001. We measured loss using the binary cross-entropy function and included two callbacks to optimize training behavior: first, we applied a learning rate stepdown function which halved the learning rate if validation loss did not improve for 5 consecutive epochs, and second, we saved the CNN configuration only after epochs in which validation loss improved, which helped to prevent overfitting to the training set.

We assessed the performance of the CNN by having it generate predictions on a test set of ca. 130,000 spectrograms whose labels were known. The test set was compiled from data from 2017 and 2018 and had no overlap with data used to train the model. The test dataset included 5,000 images containing NSO four-note location calls, 5,000 images containing barred owl eight-note calls, and 4,000 images containing barred owl single-note "inspection" calls, which we included as a separate class to reduce confusion between vocalizations of the two species. We assessed model performance using class-specific measures of precision and recall, two commonly used metrics in machine learning. Precision, or specificity, is defined as the proportion of apparent "hits" that correspond to real detections for a given class. Precision is calculated as [True Positives] / [True Positives + False Positives]. Recall, or sensitivity, represents the proportion of actual calls in the dataset that are detected and correctly labeled by the model. Recall is calculated as [True Positives] / [True Positives + False Negatives].

Output from the CNN was validated by trained human reviewers in order to generate species encounter histories to be used in occupancy models and other analyses. Currently we are validating all clips that scored higher than 0.25 for NSO, regardless of what the highest-scoring class was; all clips that scored higher than 0.95 for other target species; and a random sample of 0.5% of all other clips. These rules are intended to maximize detection power for NSOs, allow us to quickly produce encounter histories for all target species at the hexagon level, and minimize time spent on false positives. Validating a random sample of clips that did not otherwise merit consideration allows us to estimate the number of false negatives, i.e. the number of real vocalizations that would otherwise be missed under this validation scheme.