

Articles

Let's Agree to Disagree: Comparing Auto-Acoustic Identification Programs for Northeastern Bats

Tomás Nocera, W. Mark Ford,* Alexander Silvis, Christopher A. Dobony

T. Nocera

Department of Fish and Wildlife Conservation, 106 Cheatham Hall, Virginia Tech, Blacksburg, Virginia 24061

W.M. Ford

U.S. Geological Survey, Virginia Cooperative Fish and Wildlife Research Unit, Virginia Tech, 106 Cheatham Hall, Blacksburg, Virginia 24061

A. Silvis

West Virginia Division of Natural Resources, 738 Ward Road, Elkins, West Virginia 26241

C.A. Dobony

Fort Drum Military Installation, Directorate of Public Works, Natural Resource Branch, T-4848 Delahanty Avenue, Fort Drum, New York 13602

Abstract

With the declines in abundance and changing distribution of white-nose syndrome-affected bat species, increased reliance on acoustic monitoring is now the new “normal.” As such, the ability to accurately identify individual bat species with acoustic identification programs has become increasingly important. We assessed rates of disagreement between the three U.S. Fish and Wildlife Service-approved acoustic identification software programs (Kaleidoscope Pro 4.2.0, Echoclass 3.1, and Bat Call Identification 2.7d) and manual visual identification using acoustic data collected during summers from 2003 to 2017 at Fort Drum, New York. We assessed the percentage of agreement between programs through pairwise comparisons on a total nightly count level, individual file level (e.g., individual echolocation pass call file), and grouped maximum likelihood estimate level (e.g., probability values that a species is misclassified as present when in fact it is absent) using preplanned contrasts, Akaike Information Criterion, and annual confusion matrices. Interprogram agreement on an individual file level was low, as measured by Cohen's Kappa (0.2–0.6). However, site-night level pairwise comparative analysis indicated that program agreement was higher (40–90%) using single season occupancy metrics. In comparing analytical outcomes of our different datasets (i.e., how comparable programs and visual identification are regarding the relationship between environmental conditions and bat activity), we determined high levels of congruency in the relative rankings of the model as well as the relative level of support for each individual model. This indicated that among individual software packages, when analyzing bat calls, there was consistent ecological inference beyond the file-by-file level at the scales used by managers. Depending on objectives, we believe our results can help users choose automated software and maximum likelihood estimate thresholds more appropriate for their needs and allow for better cross-comparison of studies using different automated acoustic software.

Keywords: acoustic monitoring; automated acoustic identification software programs; white-noise syndrome; *Myotis lucifugus*; *Myotis serptrionalis*; *Myotis sodalis*; Anabat

Received: October 2, 2018; Accepted: June 3, 2019; Published Online Early: July 2019; Published: December 2019

Citation: Nocera T, Ford WM, Silvis A, Dobony CA. 2019. Let's agree to disagree: comparing auto-acoustic identification programs for northeastern bats. *Journal of Fish and Wildlife Management* 10(2):346-361; e1944-687X. <https://doi.org/10.3996/102018-JFWM-090>

Copyright: All material appearing in the *Journal of Fish and Wildlife Management* is in the public domain and may be reproduced or copied without permission unless specifically noted with the copyright symbol ©. Citation of the source, as given above, is requested.

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

* Corresponding author: wmford@vt.edu



Introduction

The first documentation of the fungal pathogen *Pseudogymnoascus destructans*, the causative agent of white-nose syndrome (WNS), in the United States was on February 16, 2006, at Howe's Caverns, located 40 mi (64 km) west of Albany, New York (Blehert et al. 2009). Subsequently, the pathogen moved rapidly throughout the northeastern and central regions of the United States as well as southern Canada. By 2016, WNS had expanded into Washington state (U.S. Geological Service National Wildlife Health Center 2016), and as of 2018, either presence of the fungi or actual WNS has been reported in Iowa, Kansas, Mississippi, Texas, and Wyoming (U.S. Geological Service National Wildlife Health Center 2016). The fungus grows on the epithelial tissues of hibernating bats and causes abnormal frequent arousal and activity throughout winter that consequently leads to premature loss of critical fat reserves and disruption of water balance in infected individuals (Cryan et al. 2010; Meteyer et al. 2012). To date, millions of cave-hibernating bats have died after being infected by *P. destructans* (U.S. Fish and Wildlife Service [USFWS] 2018), and populations of some myotids (*Myotis* spp.) suffered >90% reductions (Blehert et al. 2009; Frick et al. 2010).

As the effects of WNS increase, there is a growing need for techniques to accurately monitor bat population declines and document residual bat distributions (Langwig et al. 2012). The ability to record and accurately identify individual bat species is a key tool for managers tasked with conservation of WNS-impacted species. For managers, understanding the distributions and habitat associations of threatened bat species can help guide management activities and reduce potential additive stressors, such as logging practices or recreational spelunking. Pre-WNS, mist netting was widely used to sample bat distribution and abundance. However, low densities of many bat species post-WNS have led to greatly reduced capture rates, complicating documentation of true status (Coleman et al. 2014a). Accordingly, parameter estimates produced using mist-netting techniques may not be representative of the overall changes in population and activity trends and may no longer be a viable technique in WNS-impacted areas to describe current local and regional distribution of bat species.

Acoustic monitoring techniques have been widely used to assess species presence and probable absence, activity patterns among different habitats, and spatial and temporal trends both pre- and post-WNS (Sherwin et al. 2000; Britzke et al. 2002; Weller and Zabel 2002; Ford et al. 2011; Rodhouse et al. 2011; Coleman et al. 2014b). Murray et al. (2001) and Britzke et al. (2002) demonstrated acoustic sampling is generally more effective for characterizing bat community composition than mist netting in the eastern United States. Acoustic monitoring allows managers to sample larger areas and account for greater bat species richness with less effort over time than traditional mist netting (Murray et al. 1999; Coleman et al. 2014c). The benefits of acoustic sampling include 1) increased sampling nights (effort) and survey extent (sites), 2) cost efficiency over expanded mist-netting

efforts, and 3) provision of guides for focal mist netting when bat capture for foraging and day-roost radiotelemetry research or tissue sample collection is needed (Britzke et al. 2002; Coleman et al. 2014b, 2014c). Irrespective of WNS, the flexibility of sampling a larger area and a greater number of sites, encompassing a variety of habitats in less time, often makes acoustic sampling techniques preferable to other capture methods in certain geographic areas (Britzke et al. 2013; Coleman et al. 2014a).

In conjunction with the advances in acoustic sampling technology, qualitative development of automated acoustic bat identification software has allowed for the ability to identify voluminous numbers of bat call sequences relatively quickly (Schirmacher et al. 2007; Coleman et al. 2014c; Lemen et al. 2015). Advancements in acoustic sampling and identification software, combined with declining mist-net success in the post-WNS environment, led the USFWS to develop acoustic guidelines to survey and monitor for the endangered Indiana bat *Myotis sodalis* (Niver et al. 2014). Currently available automated bat acoustic identification software programs generate species identifications for call sequences using algorithms that process and classify quantitative measures of individual calls (e.g., frequency, slope, curvature, pulse rate). Software classification algorithms are trained using reference call libraries that consist of known identity calls. These known calls typically are recorded using captured bats that are hand released and more rarely are recorded from free-flying bats. Once individual acoustic files have been assigned a species identity, or classified as noise or unidentifiable, most automated identification software programs generate maximum likelihood estimates (MLEs) following the method of Britzke et al. (2002). The MLE values represent the probability that a species is misclassified as present when in fact it is absent, and they are calculated by comparing the number of files classified as each identified species to the known misclassification rates of those species in the classification algorithm as measured using known identified calls (Britzke et al. 2002; Niver et al. 2014).

Despite greater acceptance and use of acoustic sampling and subsequent automated processing software, acoustic monitoring is not free from constraints and biases. For example, Sherwin et al. (2000) noted that acoustic sampling is unable to address abundance of an individual species beyond an index of activity (i.e., one individual recorded numerous times vs. numerous individuals recorded single times at a site over a night). Furthermore, the impetus for recently developed automated software programs was to identify *M. sodalis* first and foremost (Britzke et al. 2013). Emphasis on *M. sodalis* identification may result in the erroneous omission of identifying the presence of other species, especially in regard to similar calling *Myotis* spp., and thereby add another source of bias to the analysis of bat echolocation to identify and assess the presence of other bat species. Constraints and biases in accurate species identification, whereby the program accuracy for a given species is in part tied to how well other species are identified, can



potentially lead to either constant bias (constant misidentification) or nonconstant bias (increasing misclassification under changing conditions) that may provide misleading results (Samuel et al. 1992). Therefore, understanding the biases of acoustic sampling and identification of echolocation calls is vital for proper interpretation of study results and comparison of these results among studies when different automated identification programs are used (Sherwin et al. 2000; Adams et al. 2012; Britzke et al. 2013)

Call libraries used for software algorithm training typically only incorporate high-quality search phase echolocation call sequences of known species identifications rather than the full array of echolocation calls bats can emit (Britzke et al. 2013; Lemen et al. 2015). Although most automated software programs acknowledge that only high-quality recordings will yield accurate species identification, field recordings invariably include numerous low-quality calls that may be incongruent from software development library reference calls (Lemen et al. 2015) and overall performance can be very poor (Rydell et al. 2017). Structural variations in echolocation calls due to Doppler shifts and attenuation, coupled with echolocation adjustments in response to the presence of vegetation clutter, insect abundance and types, water, or presence of other bat species, also create considerable variation among echolocation recordings (Britzke et al. 2013). High inherent intraspecific variation in echolocation compounded with interspecies overlap in echolocation call characteristics creates significant challenges in accurate species identification (Rydell et al. 2017). In the eastern United States, high-frequency myotis and the eastern red bat *Lasiurus borealis* have similar echolocation calls, and automated identification programs have made classification errors of omission and commission (Loeb and O'Keefe 2006; Brooks 2008; Britzke et al. 2013; Silvis et al. 2016b).

Lemen et al. (2015) observed that the level of agreement of bat species identification across four automated programs used in North America was not consistent at the file level, demonstrating low levels of agreement (40%) between software packages: Bat Call Identification [BCID], Inc., Kansas City, MO), Kaleidoscope (Wildlife Acoustic, Inc., Maynard, MA), Echoclass (U.S. Army Corps of Engineers, Vicksburg, MS), and Sonobat (Sonobat, Arcata, CA). Janos (2013) found only a 38% level of agreement between files when comparing BCID to Echoclass. The low level of agreement was attributable specifically to similar high-frequency call structure among myotis and *L. borealis* (Janos 2013). Low levels of agreements among software programs are consequential for cross-study inference and conservation planning (Russo and Voigt 2016). High rates of either false positive or false negative misidentification have conservation costs that may result in incorrect management decisions, including forgoing seasonal restrictions on forest harvesting, application of prescribed fire across many management ownerships, and use of obscurities and ordinance for military training. Based on perceived, but erroneous, presence or absence of the bat species of interest, management decisions can have deleterious

effects on a bat population. It is necessary to quantify the level of agreement between bat species identification programs for researchers or managers using one or more of these programs for identifying bats with confidence relative to their stewardship needs (Lemen et al. 2015).

In 2003, before the arrival of WNS in the United States, an extensive long-term acoustic monitoring project examining spatial and temporal bat distribution, activity, and occupancy was initiated at Fort Drum Military Installation (Fort Drum) in northwestern New York (Ford et al. 2011). After the local discovery of summer maternity activity of the endangered *M. sodalis* on Fort Drum, mist-netting efforts were added to the monitoring efforts in an attempt to capture and track bats to foraging and roost locations (Jachowski et al. 2016). Continuous acoustic monitoring, as well as actual captures between 2003 and 2018, revealed changes in patterns of acoustical activity pre- and post-WNS and effects on community composition and structure at Fort Drum (Dobony et al. 2011; Ford et al. 2011; Coleman et al. 2014b; Jachowski et al. 2014) and also provided a means of comparative analysis among acoustic sampling techniques (Coleman et al. 2014a, 2014c). To assess the relative agreement among three automated bat identification software programs currently widely used in the United States and qualitative identification by a trained biologist, we examined 15 y of bat echolocation recording data from Fort Drum. The intent of our study was to compare results between Kaleidoscope Pro 4.2.0, BCID 2.7d, Echoclass 3.1, and qualitative identification to describe variation among programs and highlight potential agreement and discrepancies in output and performance. Specifically, we sought to compare programs at the nightly total level used by managers as well as actual file-by-file agreement. In addition, using USFWS MLE thresholds for nightly total acceptance of a presence or absence of a species irrespective of actual file-by-file characterization, we examined differences among programs in occupancy measures, which are the basis for bat conservation and management decisions for threatened and endangered species. Because the true species identity of calls was unknown, we were not attempting to gauge program accuracy but rather to highlight potential discrepancies among programs and visual identification by a trained biologist.

Methods

Study site

We conducted our study at Fort Drum in northwestern New York. Situated at the intersection of the St. Lawrence-Great Lakes lowlands, the Tug Hill Plateau, and the foothills of the Adirondack Mountains, Fort Drum is a 43,750 ha U.S. Army installation that contains a variety of forest, wetland, and open habitats. The Niagara Escarpment lies 10–15 km west of Fort Drum and contains limestone (karst) caves used by bats for hibernation (Ford et al. 2011). The bat fauna of Fort Drum is represented by nine species of three echolocation groups: 1) high-frequency call (minimum frequency

> 40 kHz) including little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, *M. sodalis*, tricolored bat *Perimyotis subflavus*, eastern small-footed bat *Myotis leibii*, and *L. borealis*; 2) midrange-frequency call (between 25 and 40 kHz) including big brown bat *Eptesicus fuscus* and silver-haired bat *Lasionycteris noctivagans*; and 3) low-frequency call (maximum frequency < 25 kHz) including hoary bat *Lasiurus cinereus* (Coleman et al. 2014a).

Study design

We examined echolocation calls recorded during summer (May–August), and in some years early fall (September–November), of 2003–2017 at Fort Drum. We surveyed Fort Drum via 289 individual sites and 8,373 detector nights. All calls were recorded using Anabat II detectors, connected to a compact flash-storage zero-crossings analysis-interface module, and SD1 and SD2 units (Titley Electronics, Ballina, NSW, Australia). Echolocation data collected from 2003 to 2010, before the availability of automated software, were qualitatively identified by a single trained individual using an echolocation dichotomous key developed for northeastern U.S. bat species (Ford et al. 2011). Nights where the detector did not turn on properly or shut off within 2 h after sunset were excluded from the analysis. However, if a detector ran for at least 8 h, it was included in our analysis. Calls were first identified using Analook 4.7 and then were examined for call curvature values in Analyze 2.0 (Ford et al. 2011). We reidentified echolocation calls from all years using BCID 2.7d, Kaleidoscope Pro 4.2.0, and Echoclass 3.1. For each year, we visually examined two nights per site after 2010 to ensure recording accuracy and completeness. We specifically selected the nine extant bat species that occur at Fort Drum for analysis by Kaleidoscope and BCID and Echoclass' Northeast bat assemblage that completely represented the bat species present at Fort Drum. Both Kaleidoscope and BCID allow users to adjust for sensitivity and specificity; however, although Kaleidoscope does this at the MLE level, Echoclass does not allow the end user to adjust for any parameters. We adjusted the signal parameters of Kaleidoscope at the neutral (0) setting and BCID at the minimum discriminant probability of 0.35 to follow current USFWS (2018) guidelines. We analyzed agreement rates between classifiers (i.e., software programs and trained biologist) across four grouping metrics: nightly total counts, individual files, single species occupancy metrics, and environmental variables.

Total nightly echolocation passes

We used a generalized linear mixed model with a negative binomial distribution to examine general agreement from 2003 to 2010 in total nightly counts (i.e., each individual echolocation call that is identified to a species on any given night) among automated identification software and qualitative identification. We

used relative activity by individual species at the site-night level as our response variable blocked by year, holding site as a random effect, and having our treatment groups and the total number of files recorded at each site night as fixed effects. When parameter significance was indicated at $\alpha \leq 0.05$, we performed post hoc type III test comparisons of treatments (programs and qualitative identification) by comparing least-square mean estimates for each treatment to determine differences. We fit our generalized linear mixed model in SAS 9.4 (PROC GLIMMIX; SAS Inc., Cary, NC) using a negative binomial distribution.

Individual echolocation call file

To examine the possibility that agreement levels would change at a finer scale (i.e., individual call file), we further analyzed file-by-file agreement rates between two programs, Kaleidoscope Pro 4.2.0 and Echoclass 3.1, for the full range of Fort Drum data (2003–2017). We omitted qualitative identification from these analyses because individual file identifications beyond site-night totals were not completed for the later years (2011–2017). We also omitted BCID from these analyses because MLE values were not fully “equivalent” in terms of site night to those of other software. We developed a set of confusion matrices that assessed the degree of misclassification between Kaleidoscope and Echoclass at the individual file level. Our comparisons included all bat species an individual echolocation file could have been identified as from the individual program, including simply “bat” or no identification and noise (i.e., nonbat). Although confusion matrices typically compare a predicted value with a reference value (truth), we were limited in only knowing predicted identifications from classified acoustic data. Thus, we used the premise of predicted and reference values to display how one program file level identification compared to another program and then we reversed the programs' starting position so that, in turn, each program was either the predicted or reference values for each comparison. This allowed us to examine the percentage of agreement of one program with the other and to assess the proportion of disagreement (type I and type II errors) for all files identified by the reference program. We treated years (2003–2017) and the files associated with each year as independent observations. This permitted us to analyze potential shifts in a program's ability to accurately identify species through changes in bat community structure and a species' overall abundance. Although we examined each year individually, our final matrix was cumulative, encompassing all files between 2003 and 2017. We used statistical software program R to create 11×11 confusion matrices where we assigned each acoustic software program as the predictor and reference value with the function confusion matrix using the caret package (Kuhn et al. 2018). To determine the degree of program agreement, we used Cohen's Kappa (Allouche et al. 2006) in addition to calculations of



sensitivity and specificity rates as measures of true positive and true negative performance, respectively. To visualize the confusion matrix results, we plotted the output as a heatmap in R using package ggplot2 (Wicham and Chang 2016). In addition, because USFWS guidelines for acoustic surveys rely on acceptance of presence or absence of target threatened or endangered species (i.e., *M. septentrionalis* and *M. sodalis*), we repeated these confusion matrix analyses based on recalculations of species presence on a night level with MLE values (confidence score) at $\alpha = 0.05$ as a threshold. Accordingly, we grouped MLE values, regardless of program into two categories: above $\alpha = 0.05$ (considered absent) or below $\alpha = 0.05$ (high confidence as present).

Modeling environmental conditions

To assess how disagreement among software and the trained biologist classification may impact analytical outcomes, we modeled relative nightly activity of *M. lucifugus*, *M. septentrionalis*, and *M. sodalis* identified by each program and the biologist using a set of candidate generalized linear mixed models with a negative binomial distribution (Fournier et al. 2012) from our 2003–2010 data. Because our intent was to compare analytical outcomes of our different datasets rather than model habitat associations, the candidate model set represented simple hypothesis regarding the relationship between environmental conditions and bat activity. Specific environmental conditions assessed included elevation, canopy cover, land cover type, distance to road, and distance to water. We determined the best supported model for each dataset using Akaike's Information Criterion corrected for small sample size (Burnham and Anderson 2002, 2004). Across datasets, we compared the relative model rankings and support for the best supported model (weighted Akaike's Information Criterion) as well as covariate β estimates of the best supported models. We fit generalized linear mixed models in program R 3.5.1 (R Development Core Team 2018).

Results

Field collection

From 2003 to 2010, we sampled 239 total detector nights (Table 1). File identification, by species, varied by program and qualitative identification (Table S1, *Supplemental Material*). From 2011 to 2017, we sampled an additional 8,134 total detector nights. Over the entire study duration, we recorded 1,022,188 individual files of which >450,000 were identified to bat species, although the individual species totals were variable by program and qualitative identification (Tables S1 and S2, *Supplemental Material*).

Total nightly echolocation passes

Overall, there were numerous significant differences among nightly counts across bat species among the

Table 1. Minimum, maximum, and mean number of detector nights ($n = 8,373$) and total number of site locations ($n = 289$) in Fort Drum Military Installation, New York, 2003–2017.

Year	Minimum no. of nights/site	Maximum no. of nights/site	Mean no. of nights/site	Total no. of sites
2003	1	1	1.0	5
2004	1	3	1.8	17
2005	1	2	1.8	21
2006	2	2	2.0	18
2007	2	2	2.0	15
2008	1	2	1.8	12
2009	2	2	2.0	13
2010	1	6	2.6	19
2011	1	128	17.8	57
2012	1	223	38.8	55
2013	2	187	29.6	49
2014	81	131	103.6	5
2015	2	118	10.9	129
2016	8	28	14.9	81
2017	64	83	77.8	5

three programs and the echolocation passes identified visually (Table 2). For myotids, qualitative identification generally had the highest count estimate compared with acoustic software, often significantly different from at least one program (Table 2). No bat species had a full agreement between programs and the biologist. However, for each species, except *M. leibii*, there was agreement between at least two treatments (Table 2).

Individual echolocation call file and single species occupancy metric

At the individual file level and totaled night level, agreement by species between Echoclass and Kaleidoscope varied across years measured by Cohen's Kappa, ranging from 0.25 to 0.55 depending on the year, with an average Cohen's Kappa (all years and species combined) of 0.368 (Table 3). Overall agreement proportion, when either program was used as the reference, was variable among species. When Echoclass was the reference, agreement rates for *M. septentrionalis*, *M. sodalis*, and *L. borealis* were $\leq 30\%$ (Table 4). Although the majority of disagreement between myotids were intragenus or no identification, *L. borealis* was misclassified frequently as *M. lucifugus* (Table 4). When Kaleidoscope was the reference, agreement rates for all myotids were always $\leq 30\%$, with the majority of calls misclassified as *L. borealis* (Table 4). Conversely, Kaleidoscope agreed with Echoclass up to 66% of comparisons for *Lasionycteris noctivagans*, whereas Echoclass agreed with Kaleidoscope up to 57% for *L. borealis*. Irrespective of year, nightly MLE comparisons between paired program comparisons had an overall agreement, measured by Cohen's Kappa, of 0.56 for *M. lucifugus*, 0.60 for *E. fuscus*, 0.47 for *L. borealis*, 0.58 for *L. cinereus*, 0.58 for *Lasionycteris noctivagans*, 0.04 for *M. leibii*, 0.25 for *M. septentrionalis*, 0.26 for *M. sodalis*, and 0.34 for *P. subflavus*. Overall agreement proportion at 0.05 MLE or

Table 2. Separation test, associated least square mean (LS mean), and type III test for fixed effect values for the biologist and each program (Bat Call Identification [BCID], Echoclass, and Kaleidoscope) across all species on Fort Drum Military Installation, New York, 2003–2010. Raw mean and SE, LS means and SE for all species and programs on a site-night level are included. With effect, F-statistic, and P value ($\alpha < 0.05$). P values $\leq \alpha = 0.05$ are significantly different. Species include big brown bat *Eptesicus fuscus*, eastern red bat *Lasiurus borealis*, hoary bat *Lasiurus cinereus*, silver-haired bat *Lasionycteris noctivagans*, eastern small-footed bat *Myotis leibii*, little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, Indiana bat *Myotis sodalis*, and tricolored bat *Perimyotis subflavus*.

	Program	Mean	SE	LS mean	Grouping ^a	SE (LS mean)	Effect ^b	F	Pr > F
<i>E. fuscus</i>	Kaleidoscope	20.32	3.10	2.9948	A	0.1776	Classifier	3.68	0.0118
	Biologist	17.93	2.55	2.848	A	0.1777	Year	13.89	0.0002
	BCID	16.4	2.87	2.7083	BA	0.1778	Total	171.36	<0.0001
	Echoclass	11.18	1.95	2.3957	B	0.1782			
<i>L. borealis</i>	Echoclass	27.81	4.74	2.5625	A	0.1085	Classifier	24.34	<0.0001
	BCID	15.53	3.34	2.0467	B	0.1099	Year	2.77	0.0965
	Biologist	11.2	2.09	1.9414	B	0.1103	Total	547.23	<0.0001
	Kaleidoscope	5.45	0.90	1.2239	C	0.1143			
<i>L. cinereus</i>	Biologist	19.35	3.75	3.6804	A	0.1766	Classifier	12.05	<0.0001
	Kaleidoscope	7.28	1.02	2.9364	B	0.1772	Year	0.52	0.4715
	Echoclass	7.81	1.14	2.8541	B	0.1773	Total	63.75	<0.0001
	BCID	5.92	0.80	2.62	B	0.1776			
<i>Lasionycteris noctivagans</i>	BCID	6.94	0.77	2.4246	A	0.1662	Classifier	9.26	<0.0001
	Kaleidoscope	4.01	0.44	2.0254	B	0.167	Year	10.19	0.0015
	Echoclass	3.84	0.42	1.924	B	0.1672	Total	19.8	<0.0001
	Biologist	3.39	0.49	1.4924	C	0.1686			
<i>M. leibii</i>	Biologist	5.1	2.23	0.6696	A	0.2771	Classifier	16.59	<0.0001
	BCID	2.46	0.99	0.03658	B	0.2844	Year	59.75	<0.0001
	Kaleidoscope	0.88	0.19	-0.5523	C	0.2945	Total	108.62	<0.0001
	Echoclass	0.31	0.07	-1.493	D	0.3197			
<i>M. lucifugus</i>	Biologist	40.85	5.51	2.6924	A	0.1395	Classifier	16.35	<0.0001
	Kaleidoscope	35.39	5.54	2.4575	A	0.1399	Year	173.67	<0.0001
	BCID	32.68	4.90	2.4236	A	0.1399	Total	423.09	<0.0001
	Echoclass	13.58	1.79	1.7053	B	0.1418			
<i>M. septentrionalis</i>	Biologist	8.88	3.92	-0.2582	A	0.3488	Classifier	9.78	<0.0001
	BCID	5.96	3.99	-1.1324	B	0.3568	Year	28.52	<0.0001
	Kaleidoscope	5.08	3.54	-1.5846	CB	0.3635	Total	57.74	<0.0001
	Echoclass	1.53	0.91	-2.2331	C	0.3773			
<i>M. sodalis</i>	Biologist	14.97	3.47	1.3876	A	0.2179	Classifier	2.99	0.03
	BCID	16.55	3.81	1.2349	A	0.2187	Year	180.1	<0.0001
	Echoclass	10.84	2.62	1.0686	BA	0.2197	Total	209.75	<0.0001
	Kaleidoscope	14.88	3.96	0.727	B	0.222			
<i>P. subflavus</i>	Kaleidoscope	5.29	1.04	1.6164	A	0.3034	Classifier	3.68	0.0119
	BCID	2.86	0.73	1.0742	BA	0.3054	Year	8.43	0.0038
	Biologist	2.14	0.60	0.8182	B	0.3068	Total	118.45	<0.0001
	Echoclass	1.983	0.51	0.6035	B	0.3082			

^a LS means with the same grouping variable (letter) are not significantly different from each other ($\alpha = 0.05$).

^b Numerator and denominator *df* were equal among species and effect (numerator *df* classifier = 3, year = 1, total = 1; denominator *df* = 949).

less, when either Echoclass or Kaleidoscope was the reference, was variable among species and across programs (Table 5). In addition, overall agreement rates were higher at the MLE grouping level than at an individual file-by-file level (Table 6). Specifically, when Echoclass was the reference, we observed a 2.8-fold increase in agreement between *L. borealis* file-by-file level comparison and MLE grouping (Table 6). When Kaleidoscope was the reference, we observed a 4.5-fold increase in agreement between *M. lucifugus* file-by-file level comparison and MLE grouping (Table 6).

Modeling environmental conditions

For each program and the trained biologist, the null model was outperformed by all other models across all

species analyzed with regard to relative activity (Table 7). Competing models were the same between programs and the trained biologist (Table 7). The relative ranking and the relative level of support for each individual model per program and the trained biologist indicate that there is no difference between programs and the trained biologist. For *M. lucifugus*, the model with the highest relative ranking and level of support for the trained biologist was the model with three parameters (Table 7). The global model and the model with six parameters for *M. lucifugus* activity for each program had uniform relative ranking and level of support (Table 7). For *M. septentrionalis* and *M. sodalis*, the model with the highest relative ranking and level of support for the trained biologist was our global model (Table 7). For *M. septentrionalis*, relative ranking and level of support was

Table 3. Overall agreement, measured by Kappa statistic, between Echoclass and Kaleidoscope in identifying individual files on Fort Drum Military Installation, New York, 2003–2017. Kappa statistic is a metric of observed accuracy vs. expected accuracy and measures the overall agreement for all files and species.

Year	Kappa
2003	0.25
2004	0.41
2005	0.41
2006	0.41
2007	0.38
2008	0.35
2009	0.36
2010	0.55
2011	0.31
2012	0.38
2013	0.33
2014	0.42
2015	0.45
2016	0.46
2017	0.50

highest for the model with three parameters for each program (Table 7). All three competing models for *M. sodalis* activity for each program had uniform relative ranking and level of support (Table 7). Depending on the site covariate, β estimates for site covariates among treatment models and across species were either not significantly different from zero or not significantly different from each other, except for emergent wetland vegetation and grass for *M. septentrionalis* (Figure 1).

Discussion

Species recognition using acoustics has been used for many taxa (Chesmore 2004) including insects (Chesmore and Nellenbach 2001), amphibians (Acevedo and Villanueva-Rivera 2006; Han et al. 2011; Xie et al. 2018), birds (Acevedo and Villanueva-Rivera 2006; Tyagi et al. 2006; Venier et al. 2012), marine mammals (Parijs et al. 2002; Johnson et al. 2009b), and bats (Britzke et al. 2013; Janos 2013; Coleman et al. 2014c). Advancements in technol-

Table 4. Confusion matrix table representing Kaleidoscope percent agreement with Echoclass (top) and Echoclass percent agreement with Kaleidoscope (bottom) for each species in Fort Drum Military Installation, New York, 2003–2017. The diagonal of the matrix represents proportion of agreement between the two programs; anything off the diagonal represents the proportion of times Kaleidoscope disagreed with Echoclass (top) and the proportion of times Echoclass disagreed with Kaleidoscope (bottom) for a given species (indicating where most of the disagreement is occurring). Species include big brown bat *Eptesicus fuscus*, eastern red bat *Lasiurus borealis*, hoary bat *Lasiurus cinereus*, silver-haired bat *Lasionycteris noctivagans*, eastern small-footed bat *Myotis leibii*, little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, Indiana bat *Myotis sodalis*, and tricolored bat *Perimyotis subflavus*.

	Echoclass										
	<i>E. fuscus</i>	<i>L. borealis</i>	<i>L. cinereus</i>	<i>Lasionycteris noctivagans</i>	<i>M. leibii</i>	<i>M. lucifugus</i>	<i>M. septentrionalis</i>	<i>M. sodalis</i>	NoID	Noise	<i>P. subflavus</i>
Kaleidoscope											
<i>E. fuscus</i>	60	11	2	7	2	3	6	2	9	1	2
<i>L. borealis</i>	1	24	—	—	15	3	2	3	3	—	1
<i>L. cinereus</i>	3	1	69	5	1	1	1	1	13	5	4
<i>Lasionycteris noctivagans</i>	16	3	13	66	—	1	—	1	11	3	3
<i>M. leibii</i>	—	—	—	—	5	—	1	—	—	—	—
<i>M. lucifugus</i>	3	23	1	2	21	63	11	4	9	1	3
<i>M. septentrionalis</i>	—	1	—	—	3	—	3	—	—	—	—
<i>M. sodalis</i>	—	1	—	—	7	1	2	14	1	—	—
NoID	14	26	6	14	42	13	39	22	15	5	13
Noise	3	8	9	6	4	15	7	17	38	85	11
<i>P. subflavus</i>	—	3	—	—	1	1	1	—	1	—	63
	Kaleidoscope										
	<i>E. fuscus</i>	<i>L. borealis</i>	<i>L. cinereus</i>	<i>Lasionycteris noctivagans</i>	<i>M. leibii</i>	<i>M. lucifugus</i>	<i>M. septentrionalis</i>	<i>M. sodalis</i>	NoID	Noise	<i>P. subflavus</i>
Echoclass											
<i>E. fuscus</i>	30	1	1	6	1	1	1	1	4	—	—
<i>L. borealis</i>	11	57	1	2	19	24	2	16	14	1	24
<i>L. cinereus</i>	2	—	38	10	1	1	—	—	4	1	3
<i>Lasionycteris noctivagans</i>	3	—	1	19	1	1	1	1	3	—	1
<i>M. leibii</i>	—	—	—	—	6	—	1	1	—	—	—
<i>M. lucifugus</i>	—	1	—	—	1	12	—	3	1	—	1
<i>M. septentrionalis</i>	—	—	—	—	1	—	11	—	—	—	—
<i>M. sodalis</i>	—	1	—	—	4	6	2	27	2	—	—
NoID	48	37	36	41	63	49	55	46	46	20	49
Noise	6	2	23	22	4	6	7	5	26	77	5
<i>P. subflavus</i>	—	—	—	—	—	—	—	—	—	—	15

NoID = no identification.

Table 5. Kaleidoscope percent agreement with Echoclass (top) and Echoclass percent agreement with Kaleidoscope (bottom) on a nightly maximum likelihood estimate group level for both present ($\alpha \leq 0.05$) and absent ($\alpha > 0.05$) for each species in Fort Drum Military Installation, New York, 2003–2017. *N* is the sample size (for each species) Echoclass categorized present or absent (top) and the sample size (for each species). Kaleidoscope categorized present or absent (bottom). Species include big brown bat *Eptesicus fuscus*, eastern red bat *Lasiurus borealis*, hoary bat *Lasiurus cinereus*, silver-haired bat *Lasionycteris noctivagans*, eastern small-footed bat *Myotis leibii*, little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, Indiana bat *Myotis sodalis*, and tricolored bat *Perimyotis subflavus*.

	<i>N</i>	$\alpha \leq 0.05$, % Kaleidoscope agreed	<i>N</i>	$\alpha > 0.05$, % Kaleidoscope agreed
Echoclass				
<i>E. fuscus</i>	3,714	79.77	4,681	80.85
<i>L. borealis</i>	3,957	66.79	4,438	80.71
<i>L. cinereus</i>	2,590	90.07	5,805	75.55
<i>Lasionycteris noctivagans</i>	1,469	64.46	6,926	93.02
<i>M. leibii</i>	16	56.25	8,379	95.65
<i>M. lucifugus</i>	1,412	83.56	6,983	85.55
<i>M. septentrionalis</i>	112	47.32	8,283	97.23
<i>M. sodalis</i>	1,325	18.86	7,070	99.11
<i>P. subflavus</i>	758	43.53	7,637	93.24
	<i>N</i>	$\alpha \leq 0.05$, % Echoclass agreed	<i>N</i>	$\alpha > 0.05$, % Echoclass agreed
Kaleidoscope				
<i>E. fuscus</i>	3,859	76.78	4,536	83.44
<i>L. borealis</i>	3,499	75.55	4,896	73.16
<i>L. cinereus</i>	3,752	62.18	4,643	94.44
<i>Lasionycteris noctivagans</i>	1,430	66.22	6,965	92.5
<i>M. leibii</i>	373	2.41	8,022	99.91
<i>M. lucifugus</i>	2,189	53.9	6,206	96.26
<i>M. septentrionalis</i>	282	18.79	8,113	99.27
<i>M. sodalis</i>	313	79.87	8,082	86.69
<i>P. subflavus</i>	845	39	7,549	94.33

ogy and automated acoustic software programs and detectors have enhanced the ability to identify species by sound, often providing measures of inter- and intraspecific interactions (Chesmore 2004). Acoustic sampling can help in monitoring and identifying trends in population declines of once common species or rare species (Jaramillo-Legorreta et al. 2017; Xie et al. 2018). The constraints of automated acoustic detectors and software are uniform across taxa and include difficulty of tracking individuals moving in and out of acoustic range or becoming “lost” in a group of vocalizing animals (Johnson et al. 2009b). In addition, accuracy is dependent on the foundation of a high-quality reference library that includes variation between species and within species (Britzke et al. 2002; Scott Brandes 2008; Xie et al. 2018). Reduction of extraneous noise is also an inherent issue when trying to record and identify species and individuals (Britzke et al. 2002; Scott Brandes 2008). Controlling for these factors is necessary to reduce high levels of false positive and false negative identification (Towsey et al. 2012). Further restrictions are applied to acoustic monitoring for bats because calls are not only ultrasonic but also can share similar patterns and present considerable overlap among species, making automated acoustic identification of a bat species difficult in practice (Britzke et al. 2002, 2013).

As the effects of WNS on bats continue to cause declines in the distribution and abundance of affected bat species (Frick et al. 2010; Langwig et al. 2012), there

is an increased need to use acoustic sampling to replace mist-net surveys (Coleman et al. 2014a). Consequently, managers need to clearly understand the biases associated with software programs and the limitations on inferences that can be drawn from automated bat identification, because conservation decisions relative to bats may emanate from software use. Software packages such as Kaleidoscope Pro, Echoclass, and BCID can process acoustic data and compute a species identification for every bat call sequence, although each was specifically designed with *M. sodalis* identification as the program “driver” (Janos 2013; Niver et al. 2014; Lemen et al. 2015). The promise of automated identification software was that internal numerical quantification and statistical analysis would produce higher rates of correct identification and offer repeatability free of subjective biases associated with visual identification (Lemen et al. 2015). Although we cannot determine which program is most accurate, we observed that depending on the species, overall estimates among programs, across any given site and night, are highly variable. At minimum, comparison of these estimates allowed us to determine which programs are more or less conservative in their approach to identify and count an individual bat species.

Our inclusion of the trained biologist allowed for a direct link to past studies and traditional bat identification methods. It also provided insight to understand where, when, and how humans could systematically

Table 6. Kaleidoscope percent agreement with Echoclass (top) and Echoclass percent agreement with Kaleidoscope (bottom) on an individual file level and a nightly maximum likelihood estimate (MLE) level of presence ($\alpha \leq 0.05$) for each species in Fort Drum Military Installation, New York, 2003–2017. MLE values represent the probability that a species is misclassified as present when in fact it is absent on a given site night based on multiple species. Species include big brown bat *Eptesicus fuscus*, eastern red bat *Lasiurus borealis*, hoary bat *Lasiurus cinereus*, silver-haired bat *Lasionycteris noctivagans*, eastern small-footed bat *Myotis leibii*, little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, Indiana bat *Myotis sodalis*, and tricolored bat *Perimyotis subflavus*.

	Individual files	Nightly MLE
% Kaleidoscope agreement with Echoclass		
<i>E. fuscus</i>	60	79.77
<i>L. borealis</i>	24	66.79
<i>L. cinereus</i>	69	90.07
<i>Lasionycteris noctivagans</i>	66	64.46
<i>M. leibii</i>	5	56.25
<i>M. lucifugus</i>	63	83.56
<i>M. septentrionalis</i>	30	47.32
<i>M. sodalis</i>	14	18.86
<i>P. subflavus</i>	63	43.53
% Echoclass agreement with Kaleidoscope		
<i>E. fuscus</i>	30	76.78
<i>L. borealis</i>	57	75.55
<i>L. cinereus</i>	38	62.18
<i>Lasionycteris noctivagans</i>	19	66.22
<i>M. leibii</i>	6	2.41
<i>M. lucifugus</i>	12	53.90
<i>M. septentrionalis</i>	11	18.79
<i>M. sodalis</i>	27	79.87
<i>P. subflavus</i>	15	39.00

differ from an automated acoustic program and established a baseline to assess program discrepancies. However, our inclusion of the trained biologist has limitations in that we do not know the overall accuracy of the visually identified calls, nor are the efforts replicated by other observers. This latter issue relates to two questions of how representative our biologist is of all biologists trained to identify bat calls and what level of variation in observations, if any, our biologist had among years. Compared to the biologist, we observed variation in agreement in species identification among programs and between programs and the biologist. This corroborates the comparison of Jennings et al. (2008) between human identification and automated neural networks on identifying and assessing misclassifications of bat species. We identified four areas of variation in which agreement and disagreement can occur with bat echolocation identification: 1) programs and the trained biologist tended to agree on total bat counts by site night, 2) programs and the trained biologist tended to disagree on total bat counts by site night, 3) programs estimated greater bat calls than the trained biologist on a site night, and 4) programs estimated fewer bat calls than the trained biologist on a site night. It is worth noting that for sources 3 and 4 of variation among individual programs and visual identification, qualitative

identification served as a baseline to assess program discrepancies.

We found high levels of agreement between two programs: agreement between BCID and Kaleidoscope relative to each other and with qualitative identification, particularly with *M. lucifugus*. Conversely, Echoclass tended to identify fewer *M. lucifugus*. For Echoclass, this is due to inherent software constraints and considerations to specifically find or minimize misclassification of *M. sodalis* (E. Britzke, U.S. Army Corps of Engineers, personal communication; Britzke et al. 2002). BCID and Kaleidoscope tended to agree with each other on the number of *M. lucifugus* on any given site night. We found high levels of disagreement among all treatments for rare species in our study area, which was typically associated with low numbers of identified calls of these species. As an example, the biologist estimated more *M. leibii* than the other software programs, and programs disagreed with each other. At Fort Drum, *M. leibii* are considered uncommon, even before the advent of WNS, and throughout our study total counts across all years were low (Ford et al. 2011). Low numbers on a nightly basis contribute to program uncertainty on correct classification (Janos 2013). In terms of distinguishing *M. leibii* from other myotis bats, the biologist may have been able to parse out the subtle differences in overlapping diagnostic features, such as higher individual pulse minimum frequencies, to estimate greater *M. leibii* presence than the programs.

For *Lasionycteris noctivagans*, each program estimated greater numbers of bats than the trained biologist. This may be due to constraints resulting from the use of the echolocation dichotomous key for the Northeast used by the biologist (Ford et al. 2011), which did not provide sufficient differences in call characteristics between *Lasionycteris noctivagans* and *E. fuscus*. In the qualitative identification process, we first identified suspect *Lasionycteris noctivagans* calls using Analook 4.7 and then call curvature values in Analyze 2.0 to differentiate from *E. fuscus*, requiring the biologist to decide whether to proceed to the second visualization software program (Betts 2009; Ford et al. 2011). However, the automated programs have the diagnostic ability to identify *Lasionycteris noctivagans* directly (Britzke et al. 2011; Janos 2013). Moreover, at least before the advent of WNS, *Lasionycteris noctivagans* were perceived to be common only during spring and fall migratory periods at Fort Drum and across New York (Whitaker and Hamilton 1998; Cryan 2003; Ford et al. 2011), which may have resulted in subjective bias against *Lasionycteris noctivagans* identification by qualitative identification except in cases of highly diagnostic calls. Whereas automated acoustic software does not do this, this bias may be acceptable because it helps to reduce overestimation of activity and presence of rare species.

Finally, lower estimates for *M. septentrionalis* by programs relative to the trained biologist constituted the fourth example of variability. We suspect that this is because of the biologist's ability to use information on temporal context (i.e., proximity of *M. septentrionalis* calls in file sequence). Thus, human tendency to dismiss series



Table 7. Rankings of models predicting little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, and Indiana bat *Myotis sodalis* activity with three programs (Bat Call Identification [BCID], Echoclass, and Kaleidoscope) and the biologist visual identification at Fort Drum Military Installation, New York, summers 2003–2010. With relative rank and level of support for each individual model for each program w_i (model weight).

Model		BCID w_i	Echoclass w_i	Biologist w_i	Kaleidoscope w_i
<i>M. lucifugus</i>	Date _ Canopy Cover + Habitat + Distance to Water + Distance to Road + Year	0.55	0.68	0.11	0.36
	Date + Elevation + Canopy Cover + Habitat + Distance to Water + Distance to Road + Year	0.34	0.2	0.14	0.43
	Date + Habitat + Year	0.11	0.11	0.75	0.21
	Null	0	0	0	0
<i>M. septentrionalis</i>	Date _ Canopy Cover + Habitat + Distance to Water + Distance to Road + Year	0.06	0.06	0.25	0.14
	Date + Elevation + Canopy Cover + Habitat + Distance to Water + Distance to Road + Year	0.1	0.05	0.58	0.23
	Date + Habitat + Year	0.84	0.89	0.17	0.64
	Null	0	0	0	0
<i>M. sodalis</i>	Date _ Canopy Cover + Habitat + Distance to Water + Distance to Road + Year	0.24	0.23	0.26	0.21
	Date + Elevation + Canopy Cover + Habitat + Distance to Water + Distance to Road + Year	0.43	0.35	0.73	0.15
	Date + Habitat + Year	0.33	0.43	0.01	0.64
	Null	0	0	0	0

of calls, separated by seconds, as two different species may have resulted in a greater number of counts made by the biologist for *M. septentrionalis* in Fort Drum, as observed elsewhere (Fenton 1980). Also, although *M. septentrionalis* were exceedingly abundant at Fort Drum pre-WNS, the species' low-amplitude echolocation characteristics result in poor call quality (Ford et al. 2005) and as a result are more readily dismissed as no identification or noise by automated software, whereas the biologist was comfortable assigning a species identification.

Unexpectedly, we found that the overall agreement in bat identification between Echoclass and Kaleidoscope was variable on a yearly basis. This did not conform to our original expectation that although the programs might differ overall, the rates of difference would be constant across years in terms of disagreement. It is possible that with differences in underlying use of misclassification rates set by the program's development with training data, and subsequent validation, that misclassification rates between species comparisons might be influenced differently by the proportion or total amount of calls analyzed (Britzke et al. 2002; Coleman et al. 2014c). File quantity and quality used in program development affect species' detection rates and classification. These programs were developed to prioritize correct classification of *M. sodalis*, as the rationale that initially precipitated software development and use (Britzke et al. 2002; Wildlife Acoustics 2017). The subsequent listing and realization for the need to correctly identify *M. septentrionalis* highlight the unfortunate "Red Queen effect" between technology development and use in a highly dynamic environment (i.e.,

changes in management priorities among species and changing bat community structure post-WNS) outpace software development (Van Valen 1977; Barnett and Hansen 1996; Voelpel et al. 2005).

Our work showed the level and direction of disagreement within these programs for detecting species of interest (*M. lucifugus*, *M. septentrionalis*, and *M. sodalis*). For rare species in particular, analysis of only high-quality calls can reduce the number of species identifications below reality. Such low identification rates can directly impact MLE value calculations and lead to inaccurate estimates of species presence or probable absence (Britzke et al. 2002). Because programs will differ in file identifications due to classification algorithm differences and filtering, both in terms of what files will be identified, removal of noise, and extraction of call parameters from individual calls and passes, this will also result in differences in nightly activity levels. This issue may be resolved using more accurate classification algorithms that can correctly identify species from lower quality calls and by use of higher quality recording equipment and optimized detector deployment and deployment sites.

When all years were combined, on an individual file level, our results indicate that there is high disagreement between species between Echoclass and Kaleidoscope. Beyond the probable but unknown differences among programs' reference libraries used for training and validation, the frequency and pulse rate settings within each program were not equal for the end user; for example, Echoclass does not allow users to adjust the frequency and pulse count. Austin (2017) and Hyzy et al. (2018) both noted acceptable performance in both

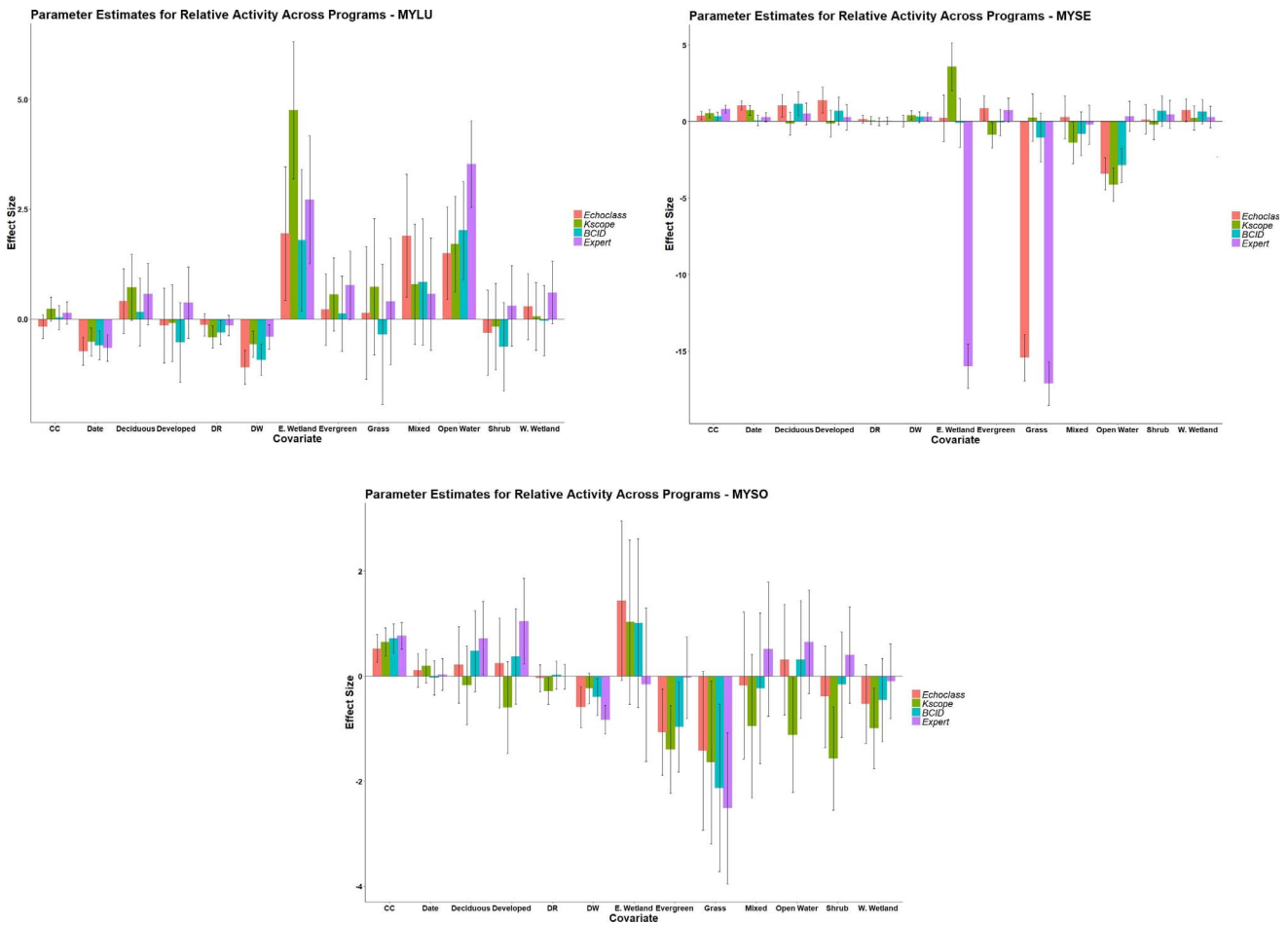


Figure 1. Parameter estimates and SE of our global negative binomial generalized linear mixed model predicting (A) little brown bat *Myotis lucifugus*, (B) *Myotis septentrionalis* northern long-eared bat, and (C) Indiana bat *Myotis sodalis* activity at Fort Drum Military Installation, New York, during summers 2003–2010. Parameters include canopy cover (CC), date, deciduous forest, developed areas, distance to road (DR), distance to water (DW), emergent wetlands, evergreen forests, grass, mixed forest, open water, shrub, and woody wetland.

programs for identifying *M. septentrionalis* presence when using the MLE threshold; however, in comparisons, file-by-file agreement generally did not exceed 40% on a site-night basis. Our findings indicate relatively higher rates of agreement between Echoclass and Kaleidoscope when using the MLE threshold either as a screen for nightly total activity or simply assessing presence for species such as *M. lucifugus*, *M. septentrionalis*, and *M. sodalis*. Currently, the MLE threshold is used by the USFWS as the determinant for *M. sodalis* presence and is subsequently necessary for sampling level of effort (Wintle et al. 2012; Niver et al. 2014). Our findings suggest that this approach for assigning presence or absence in an occupancy analysis format (MacKenzie et al. 2006) is robust. In the context of program agreement, at the MLE group level, Echoclass and Kaleidoscope provide similar results if used in the endangered species regulatory context to avoid or minimize take of *M. sodalis* or *M. septentrionalis*. In trying to develop ways to maximize consistency while minimizing uncertainties, resource management decisions necessitate that tools used to make decisions be the least biased part of the

process (Kareiva and Marvier 2011); yet, in the case of automated acoustic software this may not yet be fully possible.

The goal of controlling the level at which presence of a species is assumed (MLE values ≤ 0.05) is to minimize the rate of type I and type II errors, whereby simultaneously maximizing both sensitivity and specificity. We observed more type II errors than type I errors, which may be a factor of higher agreement of species being absent than being present. Thus, our data had higher agreement with MLE values grouped above 0.05 than below 0.05. Nevertheless, higher overall agreement rates among species using MLE grouping variables does indicate that these programs, at this level, provide some consistency in output. As the MLE threshold value is continued to be lowered (i.e., more conservative on determining species presence), this implies an increased rate of agreement. However, the cost of lowering the MLE threshold value is high, because the certainty of species presence or absence drops, causing the rate of false positive and false negative errors to increase. These errors are important to consider and acknowledge because they

can cause issues when interpreting results that are used to guide management and regulatory actions (Taylor and Gerrodette 1993; Fielding and Bell 1997). To illustrate, if misclassifications or disagreement generate high rates of type I errors (false positive), researchers and managers may take unnecessary actions that divert attention, effort, and funds away from other stewardship activities with little impact to the target species. In the context of *M. septentrionalis* and *M. sodalis*, use of forest management techniques (i.e., prescribed fire or harvesting) potentially designed to benefit targeted bats (Johnson et al. 2009a; Germain et al. 2017), but which are not actually present, may come at the cost of some other biotic factor or organism also of high conservation concern (Fisher and Wilkinson 2005; Dickinson et al. 2009; Silvis et al. 2016a, 2016b). Conversely, if misclassifications or disagreements generate high rates of type II errors (false negatives), researchers and managers may unknowingly take actions that, although beneficial to other conservation concerns, are deleterious to bats. More importantly, in terms of actions that degrade, fragment, or convert habitats, such as energy extraction and delivery, highway construction, or other forms of permanent forest conversion, the failure to account for species such as *M. septentrionalis* and *M. sodalis* and mitigate appropriately could have considerable negative impacts for these species (Baerwald et al. 2009; Northrup and Wittemyer 2013).

Depending on objectives and location, we believe our results can help users choose automated software and MLE thresholds most appropriate for their needs. First, the location of a study area and the extant bat community are important considerations in program choice, because the assemblage of bat species and community composition change both across latitudinal and longitudinal gradients, which can lead to higher program misclassification rates. Thus, matching potential identification with a known species pool becomes exceedingly important (Lemen et al. 2015). In turn, MLE values are calculated based on the species pool selected. Second, in the case of generalized bat community surveys, where the intent is to document species presence in a broad sense over a large landscape, any of the three programs we examined may suffice. In this case, where a liberal approach is sufficient, the MLE threshold value may need to be adjusted to allow for some variability. However, in the case of regulatory clearance surveys, where the intent is to ascertain localized presences (i.e., *M. sodalis* or *M. septentrionalis*) with a high degree of certainty to minimize type I and type II errors, managers may opt to use a program that is at least equitable to qualitative identification and other programs, but with the ability and optionality to be conservative in identification. In this case, the MLE threshold value may need to be adjusted to be more liberal (highly probable that a species is present).

Advances in occupancy modeling (Royle and Link 2006) that incorporate false positives have been used to estimate occurrence of bats using multiple automated identification programs (Clement et al. 2014; Austin 2017). These models allow for flexibility in determining

species presence or absence by adding a third category whereby both programs state that a species is present. Although false positive occupancy models using Echo-class and Kaleidoscope still revealed low levels of agreement on presence in some instances (Austin 2017), these same models had higher agreement when used for *M. septentrionalis* identification, as found by Hyzy et al. (2018). Although variability in agreement may be a result of the quality and amount of calls, as well as location, this modeling approach generates more precise parameter estimates than single season occupancy estimates. In this context, managers could use two programs to determine a conservative but highly accurate assessment of occupancy. Approaches such as this, for listed species such as *M. septentrionalis* and *M. sodalis*, might be the best approach moving forward in the post-WNS environment.

When modeling relative activity between programs and the trained biologist to determine whether selection of examined site covariates changed relative to the program used, the relative ranking and level of support for each individual model were the same across species and treatment. Indicating that regardless of which program is used, as well as qualitative identification, activity response resulted in similar modeled patterns of bat relative activity at Fort Drum. Specifically, among *M. lucifugus*, *M. septentrionalis*, and *M. sodalis*, there was no significant difference among β estimates, excluding emergent wetlands and grass for *M. septentrionalis*, which may be due to a combination of relatively low numbers of *M. septentrionalis* and few sampling sites in these habitats. Not only were total counts by most species similar across site nights but also program selection and qualitative identification were similar in quantifying site or habitat characteristics important for determining these species' presence on the landscape at Fort Drum. Specifically, in comparing analytical outcomes of our different datasets (i.e., how comparable programs and a trained biologist are to each other regarding the relationship between environmental conditions and bat activity), we determined high levels of consistency in both the relative rankings of the model as well as the relative level of support for each individual model. Our findings indicate that although there are inherent differences in acoustic automated software algorithms, analytical outcomes representing the relationship between environmental conditions and bat activity are the same regardless of which program or method of bat identification is used. Our results show that studies using different programs are comparable and that any difference in habitat assessment results are not driven by their choice or use of a program. Whereas our means comparison, individual file level, and MLE grouping level comparison potentially are contributory for both research and regulatory work (Niver et al. 2014), from a manager's perspective, knowing that any program or potential trained biologist can predict activity across the landscape similarly with largely congruent results is valuable for planning and implementing acoustic monitoring work.



Conclusions

Although we do not know the true accuracy of bat echolocation data we analyzed and cannot assess automated software identification accuracy directly, we did determine that the level of agreement among all species, programs, and years is variable and not wholly consistent, corroborating the results of other studies (Jennings et al. 2008; Janos 2013; Lemen et al. 2015). Nonetheless, after accounting for biases at the individual file level and grouped MLE threshold levels, the total nightly counts appear to have an acceptable amount of congruence. Specifically, in comparing analytical outcomes of our different datasets (i.e., how comparable programs and a trained biologist are to each other regarding the relationship between environmental conditions and bat activity), we determined high levels of consistency in both the relative rankings of the model as well as the relative level of support for each individual model. Improvement in the performance of automated software by incorporating the widest array of training data and expanding reference call libraries and by parametrizing how programs classify wild recordings is needed to address variation for better cross-study comparisons. Accordingly, we suggest that researchers and managers carefully consider the purposes, and setting, for which automated bat identification software is to be used relative to their monitoring needs as noted by others (Russo and Voigt 2016; Rydell et al. 2017).

Supplemental Material

Please note: The *Journal of Fish and Wildlife Management* is not responsible for the content or functionality of any supplemental material. Queries should be directed to the corresponding author for the article.

Table S1. Total nightly echolocation passes per species by program (BCID, Echoclass, and Kaleidoscope) and the biologist in Fort Drum Military Installation, New York, 2003–2010. Species include big brown bat *Eptesicus fuscus*, eastern red bat *Lasiurus borealis*, hoary bat *Lasiurus cinereus*, silver-haired bat *Lasionycteris noctivagans*, eastern small-footed bat *Myotis leibii*, little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, Indiana bat *Myotis sodalis*, and tricolored bat *Perimyotis subflavus*. BCID = Bat Call Identification.

Found at DOI: <https://doi.org/10.3996/102018-JFWM-090.S1> (25 KB DOCX).

Table S2. Total nightly echolocation passes per species for Echoclass and Kaleidoscope in Fort Drum Military Installation, New York, 2003–2017. Species include big brown bat *Eptesicus fuscus*, eastern red bat *Lasiurus borealis*, hoary bat *Lasiurus cinereus*, silver-haired bat *Lasionycteris noctivagans*, eastern small-footed bat *Myotis leibii*, little brown bat *Myotis lucifugus*, northern long-eared bat *Myotis septentrionalis*, Indiana bat *Myotis sodalis*, and tricolored bat *Perimyotis subflavus*.

Found at DOI: <https://doi.org/10.3996/102018-JFWM-090.S1> (25 KB DOCX).

Reference S1. [USFWS] U.S. Fish and Wildlife Service. 2018. Indiana bat summer survey guidance: automated acoustic bat ID software programs. Bloomington, Minnesota: U.S. Fish and Wildlife Service Midwest Region.

Found at DOI: <https://doi.org/10.3996/102018-JFWM-090.S2> (301 KB PDF); also available at <https://www.fws.gov/midwest/endangered/mammals/ina/inbasummersurveyguidance.html>.

Acknowledgments

The Fort Drum Military Installation through U.S. Army Corps of Engineers contract W9126G-15-2-0005 via the Southern Appalachia Cooperative Ecosystems Study Unit Program supported this work. We thank J. Rodrigue and C. Whitman for field assistance. Earlier drafts of this article were reviewed by B. Carstensen. The comments of the Associate Editor and three anonymous reviewers substantially improved this article.

Any use of trade, product, website, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Acevedo MA, Villanueva-Rivera LJ. 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* 34:211–214.
- Adams AM, Jantzen MK, Hamilton RM, Fenton MB. 2012. Do you hear what I hear? Implications of detector selection for acoustic monitoring of bats. *Methods in Ecology and Evolution* 3:992–998.
- Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43:1223–1232.
- Austin LV. 2017. Impacts of fire on bats in the central Appalachians. Master's thesis. Blacksburg: Virginia Polytechnic Institute and State University.
- Baerwald EF, Edworthy J, Holder M, Barclay RMR. 2009. A large-scale mitigation experiment to reduce bat fatalities at wind energy facilities. *Journal of Wildlife Management* 73:1077–1081.
- Barnett WP, Hansen MT. 1996. The red queen in organizational evolution. *Strategic Management Journal* 17:139–157.
- Betts BJ. 2009. The effect of a fuels-reduction silviculture treatment on bat activity in northeastern Oregon. *Northwestern Naturalist* 90:107–116.
- Blehert DS, Hicks AC, Behr M, Meteyer CU, Berlowski-Zier BM, Buckles EL, Coleman JTH, Darling SR, Gargas A, Niver R, Okoniewski JC, Rudd RJ, Stone WB. 2009. Bat white-nose syndrome: an emerging fungal pathogen? *Science* 323:227–227.



- Britzke ER, Duchamp JE, Murray KL, Swihart RK, Robbins LW. 2011. Acoustic identification of bats in the eastern United States: a comparison of parametric and nonparametric methods. *Journal of Wildlife Management* 75:660–667.
- Britzke ER, Gillam EH, Murray KL. 2013. Current state of understanding of ultrasonic detectors for the study of bat ecology. *Acta Theriologica* 58:109–117.
- Britzke ER, Murray KL, Heywood JS, Robbins LW. 2002. Acoustic identification. Pages 221–225 in Kurta A, Kennedy J, editors. *The Indiana bat: biology and management of an endangered species*. Austin, Texas: Bat Conservation International.
- Brooks RT. 2008. Habitat-associated and temporal patterns of bat activity in a diverse forest landscape of southern New England, USA. *Biodiversity and Conservation* 18:529–545.
- Burnham KP, Anderson DR. 2002. *Model selection and multimodal inference: a practical information-theoretic approach*. New York, New York: Springer.
- Burnham KP, Anderson DR. 2004. Multimodal inference understanding AIC and BIC in model selection. *Sociological Methods & Research* 33:261–304.
- Chesmore D. 2004. Automated bioacoustic identification of species. *Anais da Academia Brasileira de Ciências* 76:436–440.
- Chesmore ED, Nellenbach C. 2001. Acoustic methods for the automated detection and identification of insects. *Acta Horticulturae* 562:223–231.
- Clement MJ, Rodhouse TJ, Ormsbee PC, Szweczek JM, Nichols JD. 2014. Accounting for false-positive acoustic detections of bats using occupancy models. *Journal of Applied Ecology* 51.5:1460–1467.
- Coleman LS, Ford WM, Dobony CA, Britzke ER. 2014a. Effect of passive acoustic sampling methodology on detecting bats after declines from white nose syndrome. *Journal of Ecology and the Natural Environment* 6:56–64.
- Coleman LS, Ford WM, Dobony CA, Britzke ER. 2014b. Comparison of radio-telemetric home-range analysis and acoustic detection for little brown bat habitat evaluation. *Northeastern Naturalist* 21:431–445.
- Coleman LS, Ford WM, Dobony CA, Britzke ER. 2014c. A comparison of passive and active acoustic sampling for a bat community impacted by white-nose syndrome. *Journal of Fish and Wildlife Management* 5:217–226.
- Cryan PM. 2003. Seasonal distribution of migratory tree bats (*Lasiurus* and *Lasionycteris*) in North America. *Journal of Mammalogy* 84:579–593.
- Cryan PM, Meteyer CU, Boyles JG, Blehert DS. 2010. Wing pathology of white-nose syndrome in bats suggests life-threatening disruption of physiology. *BMC Biology* 8:135. <https://doi.org/10.1186/1741-7007-8-135>
- Dickinson MB, Lacki MJ, Cox DR. 2009. Fire and the endangered Indiana bat. Pages 20–22 in *Proceedings of the 3rd fire in eastern oak forests conference*. General Technical Report NRS-P-46. Newton Square, Pennsylvania: U.S. Department of Agriculture, Forest Service, Northeastern Research Station.
- Dobony CA, Hicks AC, Langwig KE, von Linden RI, Okoniewski JC, Rainbolt RE. 2011. Little brown myotis persist despite exposure to white-nose syndrome. *Journal of Fish and Wildlife Management* 2:190–195.
- Fenton MB. 1980. Adaptiveness and ecology of echolocation in terrestrial (aerial) systems. Pages 427–446 in *Animal sonar systems*. NATO Advanced Study Institutes Series. Boston, Massachusetts: Springer.
- Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38–49.
- Fisher JT, Wilkinson L. 2005. The response of mammals to forest fire and timber harvest in the North American boreal forest. *Mammal Review* 35:51–81.
- Ford WM, Britzke ER, Dobony CA, Rodrigue JL, Johnson JB. 2011. Patterns of acoustical activity of bats prior to and following white-nose syndrome occurrence. *Journal of Fish and Wildlife Management* 2:125–134.
- Ford WM, Menzel MA, Rodrigue JL, Menzel JM, Johnson JB. 2005. Relating bat species presence to simple habitat measures in a central Appalachian forest. *Biological Conservation* 126:528–539.
- Fournier DA, Skaug HJ, Ancheta J, Ianelli J, Magnusson A, Maunder M, Nielsen A, Sibert J. 2012. AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* 27:233–249.
- Frick WF, Pollock JF, Hicks AC, Langwig KE, Reynolds DS, Turner GG, Butchkoski CM, Kunz TH. 2010. An emerging disease causes regional population collapse of a common North American bat species. *Science* 329:679–682.
- Germain MJS, Kniowski AB, Silvis A, Ford WM. 2017. Who knew? First *Myotis sodalis* (Indiana bat) maternity colony in the coastal plain of Virginia. *Northeastern Naturalist* 24:N5–N10.
- Han NC, Muniandy SV, Dayou J. 2011. Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics* 72:639–645.
- Jachowski DS, Dobony CA, Coleman LS, Ford WM, Britzke ER, Rodrigue JL. 2014. Disease and community structure: white-nose syndrome alters spatial and temporal niche partitioning in sympatric bat species. *Diversity and Distributions* 20:1002–1015.
- Jachowski DS, Rota CT, Dobony CA, Edwards JW. 2016. Seeing the forest through the trees: considering roost-site selection of the endangered Indiana bat at multiple spatial scales. *PLoS ONE* 11:e0150011. <https://doi.org/10.1371/journal.pone.0150011>
- Janos GA. 2013. Utilizing acoustic methods to identify bat species, and to assess their habitat use and perception of owls. Master's thesis. Bowling Green, Ohio: Bowling Green State University.



- Jaramillo-Legorreta A, Cardenas-Hinojosa G, Nieto-Garcia E, Rojas-Bracho L, Hoef JV, Moore J, Tregenza N, Barlow J, Gerrodette T, Thomas L, Taylor B. 2017. Passive acoustic monitoring of the decline of Mexico's critically endangered vaquita. *Conservation Biology* 31:183–191.
- Jennings N, Parsons S, Pocock MJO. 2008. Human vs. machine: identification of bat species from their echolocation calls by humans and by artificial neural networks. *Canadian Journal of Zoology* 86:371–377.
- Johnson JB, Edwards JW, Ford WM, Gates JE. 2009a. Roost tree selection by northern myotis (*Myotis septentrionalis*) maternity colonies following prescribed fire in a central Appalachian Mountains hardwood forest. *Forest Ecology and Management* 258:233–242.
- Johnson M, de Soto NA, Madsen PT. 2009b. Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. *Marine Ecology Progress Series* 395:55–73.
- Kareiva P, Marvier M. 2011. Conservation science: balancing the needs of people and nature. Pages 81–156 in *Policy, protected areas, and planning*. Greenwood Village, Colorado: Roberts and Company.
- Kuhn M, Wing J, Weston J, Williams A, Keefer C, Engelhardt A, Cooper T., Mayer Z, Kenkel B, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C, Hunt T. 2018. Caret: classification and regression training. R package version 6.0-79. Available at: <https://cran.r-project.org/web/packages/caret/caret.pdf> (September 2019).
- Langwig KE, Frick WF, Bried JT, Hicks AC, Kunz TH, Marm Kilpatrick A. 2012. Sociality, density-dependence and microclimates determine the persistence of populations suffering from a novel fungal disease, white-nose syndrome. *Ecology Letters* 15:1050–1057.
- Lemen C, Freeman PW, White JA, Andersen BR. 2015. The problem of low agreement among automated identification programs for acoustical surveys of bats. *Western North American Naturalist* 75:218–225.
- Loeb SC, O'Keefe JM. 2006. Habitat use by forest bats in South Carolina in relation to local, stand, and landscape characteristics. *Journal of Wildlife Management* 70:1210–1218.
- Meteyer CU, Barber D, Mandl JN. 2012. Pathology in euthermic bats with white nose syndrome suggests a natural manifestation of immune reconstitution inflammatory syndrome. *Virulence* 3:583–588.
- Murray KL, Britzke ER, Hadley BM, Robbins LW. 1999. Surveying bat communities: a comparison between mist nets and the Anabat II bat detector system. *Acta Chiropterologica* 1:105–112.
- Murray KL, Britzke ER, Robbins LW. 2001. Variation in search-phase calls of bats. *Journal of Mammalogy* 82:728–737.
- Niver RA, King RA, Armstrong MP, Ford WM. 2014. Methods to evaluate and develop minimum recommended summer survey effort for Indiana bats. White Paper. St. Paul, Minnesota: U.S. Fish and Wildlife Service Region 3. Available: <https://www.fws.gov/midwest/endangered/mammals/inba/surveys/pdf/WhitePaperBatSurveyEffort13Jan2014.pdf> (April 2019).
- Northrup JM, Wittemyer G. 2013. Characterising the impacts of emerging energy development on wildlife, with an eye towards mitigation. *Ecology Letters* 16:112–125.
- Parijs SMV, Smith J, Corkeron PJ. 2002. Using calls to estimate the abundance of inshore dolphins: a case study with Pacific humpback dolphins *Sousa chinensis*. *Journal of Applied Ecology* 39:853–864.
- R Development Core Team. 2018. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: <http://www.R-project.org> (September 2019).
- Rodhouse TJ, Vierling KT, Irvine KM. 2011. A practical sampling design for acoustic surveys of bats. *Journal of Wildlife Management* 75:1094–1102.
- Royle JA, Link WA. 2006. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* 87:835–841.
- Russo D, Voigt CC. 2016. The use of automated identification of bat echolocation calls in acoustic monitoring: a cautionary note for sound analysis. *Ecological Indicators* 66:598–602.
- Rydell J, Nyman S, Eklof J, Jones G, Russo D. 2017. Testing the performance of automated identification of bat echolocation calls: a request for prudence. *Ecological Indicators* 78:416–420.
- Samuel MD, Steinhorst RK, Garton EO, Unsworth JW. 1992. Estimation of wildlife population ratios incorporating survey design and visibility bias. *Journal of Wildlife Management* 56:718–725.
- Schirmacher MR, Castleberry SB, Ford WM, Miller KV. 2007. Species-specific habitat association models for bats in south-central West Virginia. *Proceedings of the Annual Conference of the Southeastern Fish and Wildlife Agencies* 61:46–53.
- Scott Brandes T. 2008. Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International* 18. Available: http://www.journals.cambridge.org/abstract_S0959270908000415 (April 2018).
- Sherwin RE, Gannon WL, Haymond S. 2000. The efficacy of acoustic techniques to infer differential use of habitat by bats. *Acta Chiropterologica* 2:145–153.
- Silvis A, Gehrt SD, Williams RA. 2016a. Effects of shelterwood harvest and prescribed fire in upland Appalachian hardwood forests on bat activity. *Forest Ecology and Management* 360:205–212. Available: <https://www.sciencedirect.com/science/article/pii/S0378112715005654> (February 2019).
- Silvis A, Perry RW, Ford WM. 2016b. Relationships of three species of white-nose syndrome-impacted bats to forest condition and management. *General Tech-*



- nical Report SRS-214. Ashville, North Carolina: U.S. Forest Service Southern Research Station.
- Taylor BL, Gerrodette T. 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conservation Biology* 7:489–500.
- Towsey M, Planitz B, Nantes A, Wimmer J, Roe P. 2012. A toolbox for animal call recognition. *Bioacoustics* 21:107–125.
- Tyagi H, Hegde RM, Murthy HA, Prabhakar A. 2006. Automatic identification of bird calls using Spectral Ensemble Average Voice Prints. Pages 1–5 in 2006 14th European Signal Processing Conference, Florence, Italy.
- [USFWS] U.S. Fish and Wildlife Service. 2018. Indiana bat summer survey guidance: automated acoustic bat ID software programs. Bloomington, Minnesota: U.S. Fish and Wildlife Service Midwest Region (see *Supplemental Material*, Reference S1).
- U.S. Geological Survey National Wildlife Health Center. 2016. White-nose syndrome surveillance. Available: https://www.nwhc.usgs.gov/disease_information/white-nose_syndrome/ (February 2017).
- Van Valen L. 1977. The red queen. *American Naturalist* 111:809–810.
- Venier LA, Holmes SB, Holborn GW, Mcilwrick KA, Brown G. 2012. Evaluation of an automated recording device for monitoring forest birds. *Wildlife Society Bulletin* 36:30–39.
- Voelpel S, Leibold M, Tekie E, von Krogh G. 2005. Escaping the red queen effect in competitive strategy. *European Management Journal* 23:37–49.
- Weller TJ, Zabel CJ. 2002. Variation in bat detections due to detector orientation in a forest. *Wildlife Society Bulletin* 30:922–930.
- Wicham H, Chang W. 2016. ggplot2: create elegant data visualisations using the grammar of graphics. R Package version 2.2.1. Available at: <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf> (September 2019).
- Whitaker JO Jr, Hamilton WJ Jr. 1998. *Mammals of the eastern United States*. Ithaca, NY: Cornell University Press.
- Xie J, Towsey M., Zhang J, Roe P. 2018. Frog call classification: a survey. *Artificial Intelligence Review* 49:375–391.