



Effects of Observer Variability in the Analysis of Landbird Audio Recordings

Natural Resource Report NPS/SECN/NRR—2015/987



ON THE COVER

Waveform view of the scolding call of Carolina Wren (*Thryothorus ludovicianus*), recorded at Cumberland Island National Seashore, 2012.

Effects of Observer Variability in the Analysis of Landbird Recordings

Natural Resource Report NPS/SECN/NRR—2015/987

Clark D. Jones¹, Michael W. Byrne², and Elizabeth Kurimo-Beechuk¹

¹University of Georgia

Warnell School of Forestry and Natural Resources

180 E. Green St.

Athens, Georgia 30602

²National Park Service

Southeast Coast Inventory and Monitoring Network

University of Georgia

135 Phoenix Rd.

Athens, Georgia 30605

July 2015

U.S. Department of the Interior

National Park Service

Natural Resource Stewardship and Science

Fort Collins, Colorado

The National Park Service, Natural Resource Stewardship and Science office in Fort Collins, Colorado, publishes a range of reports that address natural resource topics. These reports are of interest and applicability to a broad audience in the National Park Service and others in natural resource management, including scientists, conservation and environmental constituencies, and the public.

The Natural Resource Report Series is used to disseminate comprehensive information and analysis about natural resources and related topics concerning lands managed by the National Park Service. The series supports the advancement of science, informed decision-making, and the achievement of the National Park Service mission. The series also provides a forum for presenting more lengthy results that may not be accepted by publications with page limitations.

All manuscripts in the series receive the appropriate level of peer review to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and designed and published in a professional manner.

Data in this report were collected and analyzed using methods based on established, peer-reviewed protocols and were analyzed and interpreted within the guidelines of the protocols. This report received informal peer review by subject-matter experts who were not directly involved in the collection, analysis, or reporting of the data.

Views, statements, findings, conclusions, recommendations, and data in this report do not necessarily reflect views and policies of the National Park Service, U.S. Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government.

This report is available from the Southeast Coast Network website (<http://science.nature.nps.gov/im/units/secn>) and the Natural Resource Publications Management website (<http://www.nature.nps.gov/publications/nrpm/>). To receive this report in a format optimized for screen readers, please email irma@nps.gov.

Please cite this publication as:

Jones, C. D., M. W. Byrne, and E. A. Kurimo-Beechuk. 2015. Effects of observer variability in the analysis of landbird recordings. Natural Resource Report NPS/SECN/NRR—2015/987. National Park Service, Fort Collins, Colorado.

Contents

	Page
Figures.....	iv
Tables.....	v
Executive Summary	vii
Introduction.....	1
Methods.....	3
Study Area	3
Cumberland Island National Seashore (CUIS).....	3
Horseshoe Bend National Military Park (HOBE)	3
Kennesaw Mountain National Battlefield Park (KEMO).....	3
Sampling Design	4
Sampling Methodology	4
Recording Devices.....	4
Recording Evaluation	5
Data Analysis.....	6
Results.....	9
Conclusions and Recommendations	17
Literature Cited	19
Appendix A. Detection history discrepancy by park and observer.....	23
Appendix B. AOU species codes.....	29

Figures

	Page
Figure 1. Map of Southeast Coast Network park unit locations.	6
Figure 2. Occupancy estimates and 95% confidence intervals for species that met minimum sample size criteria for generating reliable estimates t CUIS.....	12
Figure 3. Occupancy estimates and 95% confidence intervals for species that met minimum sample size criteria for generating reliable estimates at HOBE	13
Figure 4. Occupancy estimates and 95% confidence intervals for species that met minimum sample size criteria for generating reliable estimates at KEMO	14
Figure 5. Power analysis for detecting a 50% decline in occupancy when initial occupancy is 0.8 and sites are visited five times.....	15
Figure 6. Power analysis for detecting a 50% decline in occupancy when initial occupancy is 0.68, and sites are visited five times.....	16

Tables

	Page
Table 1. Number of sampling locations and number of recording minutes evaluated by each observer.	5
Table 2. Total number of species detected by each observer at each park.....	10
Table 3. Jaccard index of similarity values across all observers in all parks.	10
Table 4. Species detected at a significantly different ($p \leq 0.10$) number of locations at each park.	11
Table A-1. Detection discrepancy by species and observer (1 vs. 2) in CUIS 2012.....	23
Table A-2. Detection discrepancy by species and observer (2 vs. 3) in CUIS 2012.....	24
Table A-3. Detection discrepancy by species and observer (1 vs. 3) in CUIS 2012.....	24
Table A-4. Detection discrepancy by species and observer (1 vs. 2) in HOBE 2012.	25
Table A-5. Detection discrepancy by species and observer (2 vs. 3) in HOBE 2012.	25
Table A-6. Detection discrepancy by species and observer (1 vs. 3) in HOBE 2012.	26
Table A-7. Detection discrepancy by species and observer (1 vs. 2) in KEMO 2012.	26
Table A-8. Detection discrepancy by species and observer (2 vs. 3) in KEMO 2012.	27
Table A-9. Detection discrepancy by species and observer (1 vs. 3) in KEMO 2012.	27

Executive Summary

The Southeast Coast Network uses automated recording devices (ARDs) as a tool for monitoring landbird and vocal anuran communities in the Southeast. Many advantages exist for using ARDs as a monitoring tool. First, ARDs can be deployed for long periods of time and programmed to collect data on multiple occasions without requiring any repeat visits beyond retrieving the device upon completion. Second, hazards associated with field work in remote and potentially unsafe urban areas are mitigated by not visiting sampling locations during early morning (i.e., for landbirds) or late evening (i.e., for vocal anurans). Third, unlike point-count surveys, recordings create a permanent record of the bird-monitoring event that can be evaluated and confirmed by multiple experts, while also characterizing the soundscape that can be used to address future questions.

Despite the development of computer-aided procedures or algorithms that have been shown to be successful at identifying multiple species in an automated or semi-automated process (Holmes et al. 2014) these generally require a substantial amount of time and effort to develop. Consequently, all landbird recordings collected in the Southeast Coast Inventory and Monitoring Network are currently manually evaluated (i.e., listen to playback and record all vocalizing species) for the presence of bird vocalizations until an automated or semi-automated process is finalized. Five 12-minute recording events were evaluated for each point (sampling location), temporally distributed throughout the breeding season (i.e., March–May) at each park no fewer than two days from the previously evaluated event. Sampling locations are spatially distributed at least 200 meters from one another to avoid double counting.

Because bird monitoring requires a very specific skill set, which is generally the result of several years of training and practice, the monitoring protocols define the minimum requirements that field personnel must meet to commence fieldwork or evaluate recordings. Because of the prerequisite experience necessary of field personnel, protocol training generally focuses on the methodological components and means to reduce interobserver variability. Recognizing that variability exists between observers conducting bird surveys, we sought to quantify the effect of the variation between multiple observers analyzing survey recordings by examining: (1) differences between observers, (2) differences in probability of detection, (3) differences in resulting occupancy estimates, and (4) the effects of this variation on the power to detect trends in occupancy. The overall goal of this study was to examine how observer differences in species recognition from recorded surveys might affect data interpretation for this monitoring effort within the Southeast Coast Network.

We quantified differences between observers for a subset of bird survey recordings at three parks in the Southeast Coast Network during 2012: Cumberland Island National Seashore, Horseshoe Bend National Military Park, and Kennesaw Mountain National Battlefield Park. Our results indicated the following:

- Within each park, the total number of species detected by each observer differed by as many as 12 species (Cumberland Island National Seashore) and as few as three species (Horseshoe Bend National Military Park). When examining the similarity between the species detected at

each park, through pairwise combination of observers, the Jaccard Similarity Index values ranged from 0.68–0.84.

- There were 13 species across the three parks that were detected at a significantly different number of sampling locations at each park. These species tended to be birds that do not have true “songs” (e.g., woodpeckers, nuthatches, Purple Martin, Chimney Swift, and Green Heron).
- Restricting analysis to those species which met the criteria for reliable occupancy estimates (MacKenzie and Royle 2005) reduced the mean difference between occupancy estimates by over 50%, but had little effect on average differences in probability of detection between observers.
- Given the current bird composition and distribution, methods, and sampling design, we identified a threshold whereby a 50% change in occupancy was only detectable with a power of 0.8 when initial occupancy was ≥ 0.68 , and probability of detection was > 0.65 (i.e., the most common species).
- We recommend park managers identify species or groups of species of management interest or concern (e.g., shorebirds, warblers, or wading birds) and seek to establish specific management objectives to ensure sufficient data are collected to make informed decisions.

Introduction

Landbirds were selected as one of the monitoring targets, or vital signs, during the Southeast Coast Network's vital signs selection process (DeVivo et al. 2008). The monitoring methodology implemented by the Southeast Coast Network focuses on detection and non-detection (often referred to as presence/absence) of species that occur in Southeast Coast Network parks during the breeding season (i.e., March–May; Byrne et al. 2014). Occupancy estimation is one of the primary analytical tools used with these datasets.

Interobserver variability in bird monitoring has been well documented (Cunningham et al. 1999, Diefenbach et al. 2003). This form of non-sampling error is linked to wide range of variables, such as bird vocalization behavior (e.g., rate or perching location used for vocalization), the frequency range and properties of the vocalizations and the effect of the structure of the physical environment on sound attenuation of different frequencies (i.e., higher frequency tones attenuate more rapidly than lower frequency tones and any structure along the path accelerates this pattern), and in most scenarios is a challenge to control. Regardless, error associated with observer differences can inflate abundance estimates and provide biased occupancy estimates of occupancy parameters, and incorrect trends (Sauer et al. 1994, Kendall et al. 1996, Royle and Link 2006, McClintock et al. 2010b, Miller et al. 2011).

The Southeast Coast Network uses automated recording devices (ARDs) as the primary tool for monitoring landbirds. ARDs have been used in a number of instances to monitor wildlife including nocturnal-bird monitoring in the tropics (Goyette et al. 2011), surveys for species of conservation concern, rare or potentially extinct species surveys (Fitzpatrick et al. 2005), and the development of automated detection methods (Bardeli et al. 2010; Holmes et al. 2014).

The majority of published work has revealed that automated recording techniques are at least equally effective compared to physical surveys (e.g., point counts) for monitoring bird species occurrences. (Haselmayer and Quinn 2000, Acevedo and Villanueva-Rivera 2006, Celis-Murillo et al. 2009, Venier et al. 2012). Further, skilled observers evaluating recordings detect similar numbers and species to those collected in the field (Hobson et al. 2002, Rempel et al. 2005, Holmes et al. 2014). However, Hutto and Stutzman (2009) found contrasting results in that laboratory identification was more difficult than field identification, and resulted in higher rates of misidentifications (i.e., false positives). Rempel et al. (2014) found that interobserver variability of interpretation of field recordings ranged from 0–100%, consisted of false negative and misidentification errors, and was likely associated with species characteristics (i.e., vocalization loudness and frequency).

Because ARDs focus on detection and non-detection (i.e., presence/absence), occupancy estimation is a common analytical tool (MacKenzie et al. 2002, MacKenzie et al. 2003). The influence of interobserver variability and the effect this form of non-sampling error has on occupancy and detection probability estimates, however, is an emerging effort (e.g., Royle and Link 2006, McClintock et al. 2010a). McClintock et al. (2010a, 2010b) found that even small rates of observer error (i.e., false positives, a species is incorrectly identified as present/detected) had a substantial

influence on occupancy estimates, specifically overestimating occupancy, and that unmodeled heterogeneity in detection probabilities underestimated occupancy.

Recognizing that substantial variation exists between observers conducting bird surveys, we sought to examine the repeatability of species identification across observers listening to recordings obtained from ARDs and quantify the variation between multiple observers analyzing survey recordings by examining: (1) identification discrepancies, (2) differences in probability of detection, (3) differences in occupancy estimates, and (4) the effects of this variation on the power to detect trends in occupancy. The overall goal of this study was to examine how observer differences in species recognition from recorded surveys might affect data interpretation for this monitoring effort in the Southeast Coast Network.

Methods

Study Area

The effects of observer variability in the evaluation of landbird recordings were evaluated in datasets from three parks within the Southeast Coast Inventory and Monitoring Network. These parks included Cumberland Island National Seashore (CUIS), Horseshoe Bend National Military Park (HOBE), and Kennesaw Mountain National Battlefield Park (KEMO) (Figure 1).

Cumberland Island National Seashore (CUIS)

Cumberland Island National Seashore is located on the Georgia coast between St. Andrews Sound and the mouth of the St. Marys River. At 17.5 miles (28.2 km) long, Cumberland Island National Seashore is one of the largest and most diverse islands on the Atlantic Coast. Almost half of the Seashore's 36,415 total acres (14,737 ha) is a *Spartina* spp. dominated salt marsh community. These areas also include mud flats and six tidal creeks which provide habitat for a diverse assemblage of fauna. The terrestrial areas of the Park consist of mature Virginia live oak (*Quercus virginiana*) and cabbage palm (*Sabal palmetto*) maritime forests, extensive dune systems, interdunal wetlands, freshwater systems, and mixed pine and hardwood forests in the northern part of the Park; however, vegetation communities have been significantly influenced through historic activities such as logging for shipbuilding and agriculture. Approximately 9,886 acres (4,000 ha) of the northern portion of the island are designated as a wilderness area. Cumberland Island National Seashore is considered a globally Important Bird Area by the National Audubon Society (designated in 2014) with 307 documented bird species.

Horseshoe Bend National Military Park (HOBE)

Horseshoe Bend National Military Park consists of 2,040 acres (826 ha) situated near the southern end of the Piedmont Plateau. The Park consists of low rolling hills, with elevations from 600 to 711 feet (183–217 m) above sea level, and is bisected by three-and-a-half miles (5.6 km) of the flow-regulated Tallapoosa River. Park lands were extensively cultivated from 1832 until the establishment of the Park in 1959 and are in various stages of ecological succession. In many places, pines have replaced the climax hardwoods that existed prior to conversion. The lands once used for agriculture have reverted to forests or remained as open fields. The predominant forest community type is mesic beech-oak-hickory (*Fagus* spp., *Quercus* spp., *Carya* spp.) with some loblolly pine (*Pinus taeda*). Drier areas and ridge tops are dominated by loblolly pine (*P. taeda*) with scattered longleaf pine (*P. palustris*) along the northern third of the Park. The understory is relatively open and dominated by sapling American elm (*Ulmus americana*), blueberries (*Vaccinium* spp.), Carolina silverbell (*Halesia carolina*), muscadines (*Vitus* spp.), and ferns. The Park contains many species of plants endemic to the Piedmont region, but also species associated with the Southeastern and Southern Coastal Plains. Horseshoe Bend National Military Park has 119 documented bird species (NPSpecies 2014) with many considered species of conservation concern (USFWS 2008).

Kennesaw Mountain National Battlefield Park (KEMO)

Kennesaw Mountain National Battlefield Park, a 2,923-acre (1,183-ha) park in the Atlanta metropolitan area, includes the 1,808-foot (551-m) peak of Kennesaw Mountain, Little Kennesaw Mountain, and hundreds of acres of mixed hardwood and pine forests intermixed with numerous

grassy fields. Forested areas generally consist of oaks (*Quercus spp.*), hickories (*Carya spp.*), and loblolly pine (*P. taeda*). Because of the Park's location in the metropolitan Atlanta area, it is one of the most visited parks in the National Park System and is influenced by many anthropogenic stressors (e.g., non-native species, adjacent land fragmentation and alteration, and air and water pollutants). Because Kennesaw Mountain National Battlefield Park is essentially a protected island in a large urban landscape, it is a major stopover and breeding site for many species of Neotropical and resident birds. The Park is ranked as one of the top 40 birding hotspots in the United States (Bird Watching Daily 2013) and thousands of recreational birders visit Kennesaw Mountain National Battlefield Park during spring and fall migration. The Park is known as a regular stopover site for Cerulean Warbler (Sullivan et al. 2014), a species of conservation concern throughout its range (USFWS 2008), in addition to over 30 species of Neotropical warblers. Kennesaw Mountain National Battlefield Park has 207 known bird species (NPSpecies 2014) and is recognized as a state Important Bird Area by the National Audubon Society (designated in 2014).

Sampling Design

A spatially balanced unstratified sampling design was used for each park and the park boundary was used as the sampling frame. The RRQRR algorithm, developed by Theobald et al. (2007), was used to generate a suite of sampling locations for each park. In summary, all areas within each park boundary were converted to a grid of 0.5-ha cells where the center point of each grid cell served as the sampling location. The sampling design used is detailed in Byrne et al. (2014). A minimum distance of 200 m between sampling locations was used maintain spatial independence. Thirty sampling locations were monitored at each of the parks, however because of time constraints, not all recordings from available sampling locations were evaluated for this paper.

Sampling Methodology

Recording Devices

Automated recording devices were deployed for a 77-day period during the spring at each sampling location within each park (Byrne et al. 2014). Automated recording devices were used because of their substantial safety advantages, their ability to be deployed for long periods of time, their collection of data multiple times a day without revisitation to the point, and creation a permanent record of the soundscape. The deployment period and duration was chosen to optimize peak vocalization of breeding species, accommodate variable weather patterns that influence vocalization and subsequent detection, and to produce a detection history for occupancy models. The ARD currently used by the Southeast Coast Network is a Song Meter™ manufactured by Wildlife Acoustics®, Concord, Massachusetts, USA. The device consists of a digital timer and recording device, four rechargeable D-cell batteries, and a flash memory card housed in a waterproof case with two microphones mounted on the top of the device. Automated recording devices were programmed to record two 12-minute events: one beginning at 7:30 a.m. and another at 8:00 a.m. with a pause of three days between each recording event (Byrne et al. 2014). Although only one of these recordings per day is used for analysis (i.e., the 8:00 a.m. recording), two intervals are recorded to provide a backup recording if the initial recording is not useable due to background noise (i.e., overflight noise, a train, or a barking dog). This schedule collects 20 discrete sampling events for each sampling location.

In spite of the development of computer-aided algorithms or processes which are successful at identifying multiple species in an automated process (Holmes et al. 2014), and an ongoing effort by the Southeast Coast Network team, current efforts in the Southeast Coast Network rely on manual identification. At the time of this paper, tractable methods to identify all species (i.e., the entire vocalizing bird community) recorded by automated recording devices are not widely utilized; although many are currently in development. This is due in part to the large amount of time required to develop accurate algorithms, or other processes, that facilitate automatic detection while maintaining cost- and time-effectiveness. Currently, the Southeast Coast Network manually evaluates all bird-monitoring data by listening to a subset of five 12-minute recordings collected from each sampling location in each park during the breeding season.

Recording Evaluation

Five 12-minute recordings at a subset of sampling locations from each park were evaluated by three different observers with varying levels of experience. Observer 1 had over five years of experience conducting point counts in the Southeast, Observer 2 had one field season of experience conducting point counts in the Southeast, and Observer 3 had no previous experience conducting point counts but did have some ornithological fieldwork experience. All observers were trained for a minimum of two weeks and tested prior to the initiation of recording evaluation. Each observer then listened to each recording while simultaneously examining spectrograms in the Song Scope™ software (Wildlife Acoustics®) environment (Table 1). All birds detected in the recording were documented. Each observer was allowed to use any means necessary to identify each vocalization including internet resources and other bird identification audio. Each recording was reviewed until confident all species were detected and correctly identified. If there was excessive background noise in a recording or recording device failed to produce a recording of sufficient quality due to technical issues (< 6% of all recordings), the recording was omitted prior to analysis.

Table 1. Number of sampling locations and number of recording minutes evaluated by each observer.

Park	Number of Sampling Locations Evaluated	Total Minutes of Recordings Evaluated
Cumberland Island National Seashore	20	1,092
Horseshoe Bend National Military Park	13	768
Kennesaw Mountain National Battlefield Park	21	1,188
Totals =	54	3,048

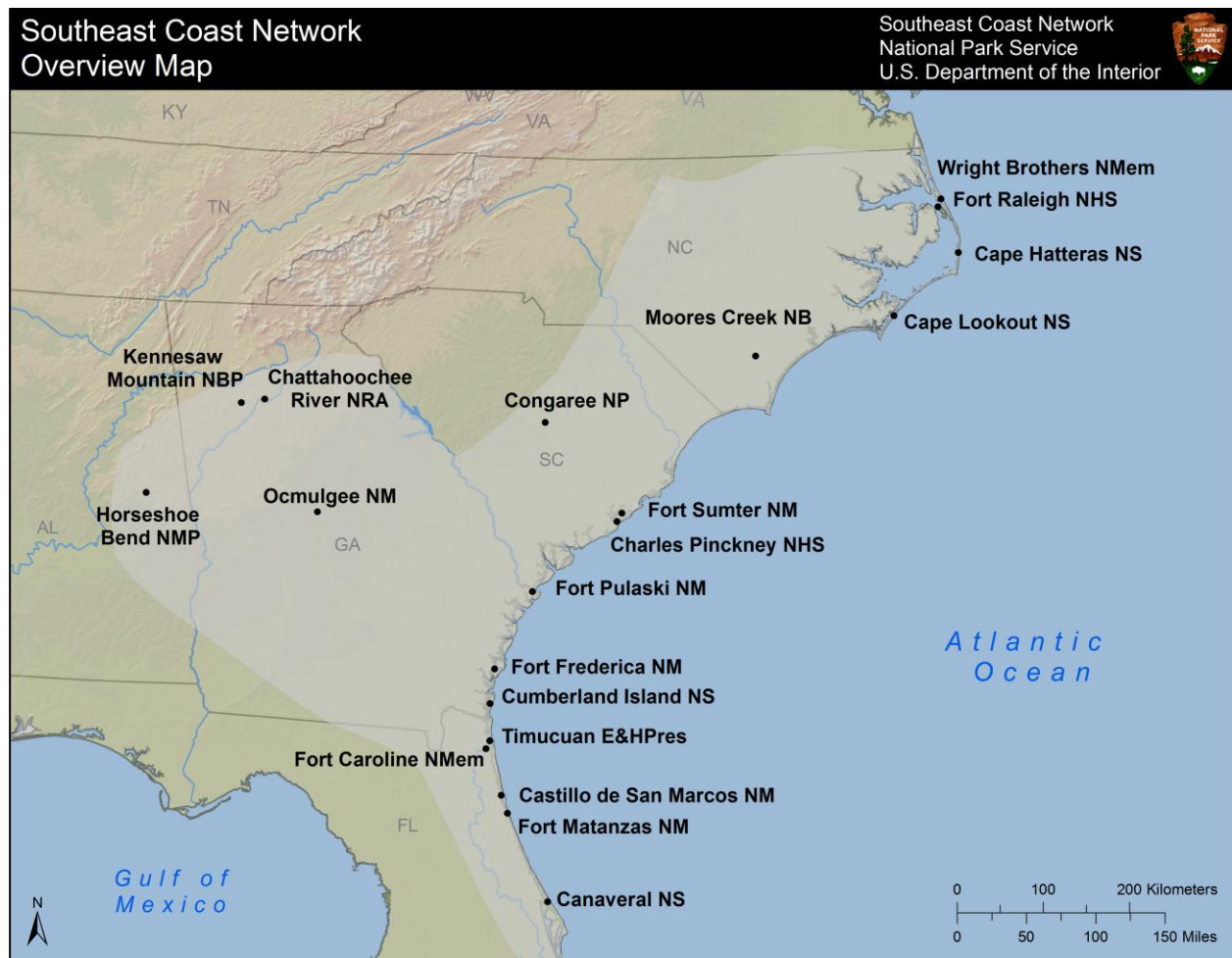


Figure 1. Map of Southeast Coast Network park unit locations.

Data Analysis

We compared species identification and resulting occupancy estimates across three observers for recordings made at three national parks in the Southeast Coast Network in 2012: Horseshoe Bend National Military Park, Kennesaw Mountain National Battlefield Park, and Cumberland Island National Seashore. Three approaches were used to examine observer differences in recording evaluation: (1) the Jaccard Similarity Index was used to compare the total similarity of species identified by each pairwise combination of observers and the encounter histories of all species that were common to each pairwise combination of observer, (2) the Fisher's Exact Test was used to test the null hypothesis that the number of sampling locations in which a species was detected did not differ between observers, and (3) Package unmarked (Fiske and Chandler 2011) was used in the R statistical environment (R Development Core Team 2014) to generate estimates of occupancy while accounting for probability of detection for all species across all observers. An average occupancy estimate was generated for each species, by park, by observer, while accounting for the probability of detection and effects of sampling date and noise. Background and ambient noise were quantified in a standardized manner following the protocol outlined in Byrne et al. (2014). We created four candidate models for each species: (1) a null detection model (i.e., constant detection), (2) a detection

model including noise as a covariate, (3) a detection model including sampling date as a covariate, and (4) a detection model including the additive effect of noise and date. We did not include any models with habitat covariates for occupancy.

MacKenzie and Royle (2005) provided general guidelines for allocating survey effort based on rates of probability of detection and occupancy in order to generate reliable estimates of occupancy. These guidelines were intended to provide a starting point for designing surveys for an estimated level of occupancy and they provided criteria for the number of surveys to conduct based on the rate of occupancy and probability of detection for a given species. Here we used these criteria to evaluate initial survey results on a subset that met the criteria. The goal was to compare how limiting the analysis to only species that met the criteria outlined by MacKenzie et al. (2005) would affect the mean differences in the occupancy estimates across observers. We predicted that species that met the criteria would have lower mean differences in occupancy estimates across observers.

Last, we conducted a power analysis to determine how the existing sampling design of the Southeast Coast Network landbird monitoring protocol (Byrne et al. 2014) is affected by differences in detectability to illustrate how differences in observer detectability may influence the monitoring program's ability to detect trends in changes of occupancy. To conduct the power analysis, we used the approach by outlined by Guillera-Aroita and Lahoz-Monfort (2012) and then simulated decreasing levels of probability of detection and varying rates of initial occupancy. The analysis generated estimates of power to detect a 50% decline in occupancy between two sampling periods.

Results

The total number of species detected by each observer differed by as many as 12 species at Cumberland Island National Seashore and as few as three species at Horseshoe Bend National Military Park (Table 2). When examining the similarity between the species detected at each park, by each pairwise combination of observers, Jaccard Similarity Index values ranged from 0.68–0.84 (Table 3).

There were 13 species across the three parks that were detected at a significantly different number of sampling locations at each park (Table 4, Appendix A). All three observers detected Acadian Flycatcher at a significantly different number of locations at Cumberland Island National Seashore, while only two of the observers differed for Acadian Flycatcher at Kennesaw Mountain National Battlefield Park. Three woodpecker species, Hairy Woodpecker, Northern Flicker, and Pileated Woodpecker, were detected at a significantly different number of sampling locations between at least two observers at both Kennesaw Mountain National Battlefield Park and Cumberland Island National Seashore. Horseshoe Bend National Military Park had only one species that was detected at a significantly different number of sampling locations between two observers: Brown-headed Nuthatch ($p = 0.07$). The p -values for both Scarlet Tanager and Summer Tanager were suggestive of a difference in detection at a different number of locations at Kennesaw Mountain National Battlefield Park and Cumberland Island National Seashore ($p = 0.11$) between at least two observers.

When examining mean differences in rates of occupancy across all observers, similar patterns were observed at each park. The mean difference in occupancy across all species and all observers at Cumberland Island National Seashore was 0.26 (SD = 0.21) and the mean difference in probability of detection across all species was 0.08 (SD = 0.10). The mean difference in occupancy across all species and all observers at Horseshoe Bend National Military Park was 0.24 (SD = 0.22) and the mean difference in probability of detection across all species was 0.07 (SD = 0.08). The mean difference in occupancy across all species and all observers at Kennesaw Mountain National Battlefield Park was 0.29 (SD = 0.21) and the mean difference in probability of detection across all species was 0.09 (SD = 0.09).

When only species which met the criteria for reliable estimates (MacKenzie and Royle 2005) are examined, the mean difference between occupancy rates decreased across all observers (Figures 2–4). For example, the mean difference for Cumberland Island National Seashore occupancy across observers dropped from 0.26 to 0.12 when species that did not meet the criteria were excluded. The mean differences at Horseshoe Bend National Military Park (0.24 versus 0.11), and Kennesaw Mountain National Battlefield Park (0.29 versus 0.18) also decreased. Some estimates, however, were still twice as high between observers, although 95% confidence intervals overlapped for most species. For example, the occupancy estimate for White-eyed Vireo from Observer 1 at Horseshoe Bend National Military Park was 0.31 (SE = 0.12), while the estimate from Observer 2 was 0.66 (SE = 0.15). When examining differences in probability of detection between observers using only species that met the criteria, the average difference increased slightly for all parks, Cumberland Island National Seashore: 0.13 (SD = 0.16) versus 0.08, Horseshoe Bend National Military Park:

0.08 (SD = 0.08) versus 0.07, and Kennesaw Mountain National Battlefield Park: 0.11 (SD = 0.11) versus 0.09.

The power analysis indicated that the current sampling design of 30 sampling locations and five visits can detect a 50% change in occupancy if the initial occupancy is at least 0.80 and the probability of detection is 0.36 (Figure 5). We also identified a threshold whereby a 50% change in occupancy was only detectable with a power of 0.8 when initial occupancy was ≥ 0.68 , and probability of detection was approximately > 0.65 (Figure 6). Across all observers, 7–9 species, depending on observer, met this threshold at Cumberland Island National Seashore, 3–4 species at Horseshoe Bend National Military Park, and 6–7 species at Kennesaw Mountain National Battlefield Park.

Table 2. Total number of species detected by each observer at each park.

Park	Number of Species		
	Observer 1	Observer 2	Observer 3
CUIS	51	39	45
HOBE	41	38	40
KEMO	61	59	53

Table 3. Jaccard index of similarity values across all observers in all parks.

Park	Jaccard Similarity Index		
	Observer 1 vs. 2	Observer 1 vs. 3	Observer 2 vs. 3
CUIS	0.73	0.68	0.75
HOBE	0.68	0.77	0.84
KEMO	0.82	0.75	0.73

Table 4. Species detected at a significantly different ($p \leq 0.10$) number of locations at each park. *P*-values from Fisher's Exact Test when the null hypothesis was rejected are presented. Suggestive *p*-values ($p=0.11$) also presented. Species not presented were not significantly different. — indicates no difference in the observer combination.

Park	Species	Obs 1 vs. Obs 2	Obs 1 vs. Obs 3	Obs 2 vs. Obs 3
CUIS	Acadian Flycatcher	0.10	0.10	0.10
CUIS	Brown-headed Cowbird	—	—	0.06
CUIS	Chimney Swift	0.00	0.00	0.00
CUIS	Green Heron	—	—	0.11
CUIS	Pileated Woodpecker	0.02	0.02	0.02
CUIS	Purple Martin	—	—	0.02
CUIS	Summer Tanager	—	—	0.11
HOBE	Brown-headed Nuthatch	0.07	—	—
KEMO	Acadian Flycatcher	—	0.13	—
KEMO	American Robin	0.06	—	0.06
KEMO	Chipping Sparrow	0.00	—	0.00
KEMO	Eastern Towhee	0.09	—	—
KEMO	Hairy Woodpecker	0.02	—	0.05
KEMO	Northern Flicker	0.06	—	0.01
KEMO	Pileated Woodpecker	0.02	—	0.06
KEMO	Scarlet Tanager	0.09	—	0.04

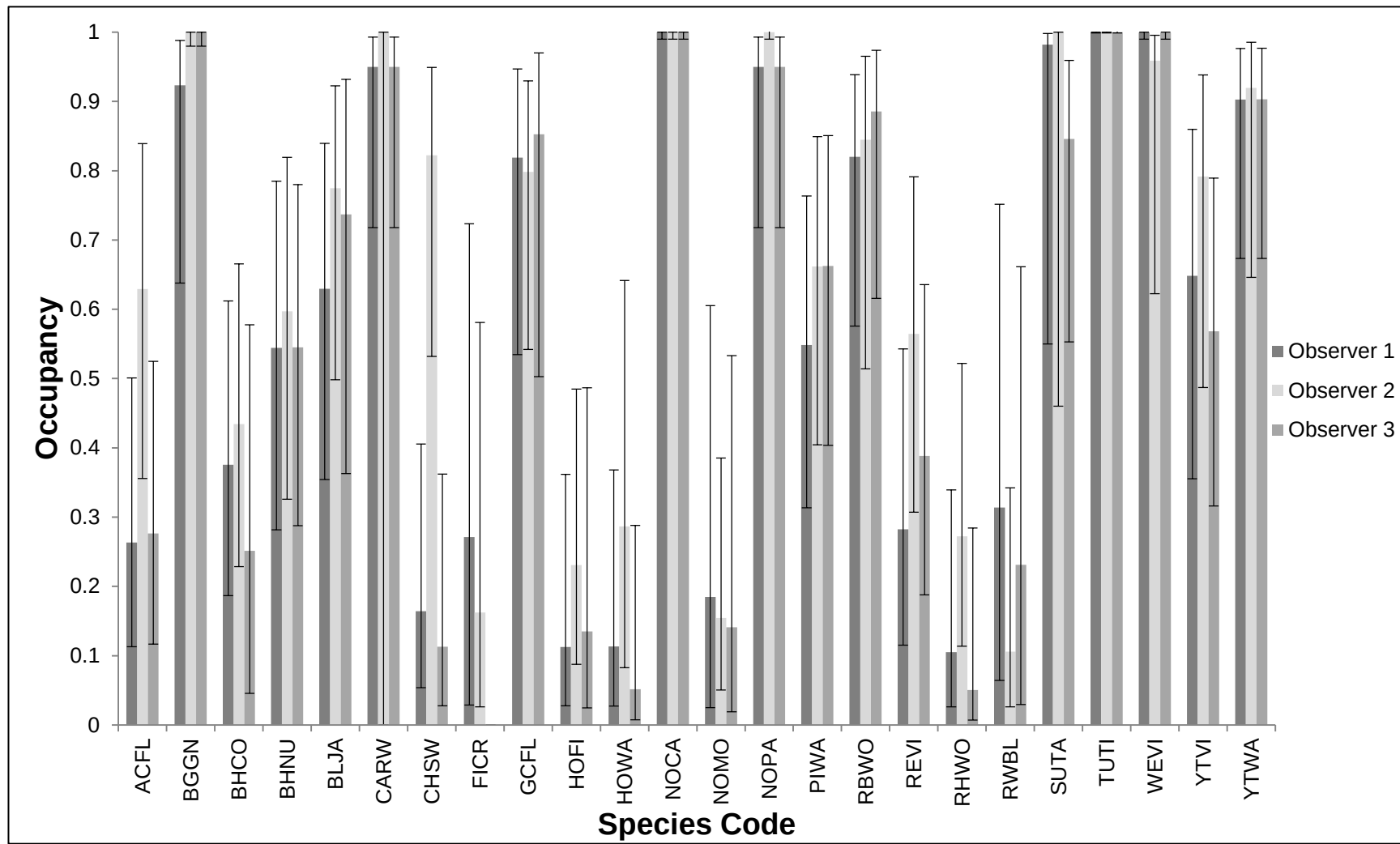


Figure 2. Occupancy estimates and 95% confidence intervals for species that met minimum sample size criteria for generating reliable estimates of occupancy (MacKenzie and Royle 2005). Results from three observers listening to identical recordings recorded at CUIS, 2012. Species codes adhere to four-letter codes from the American Ornithologists' Union (Appendix B).

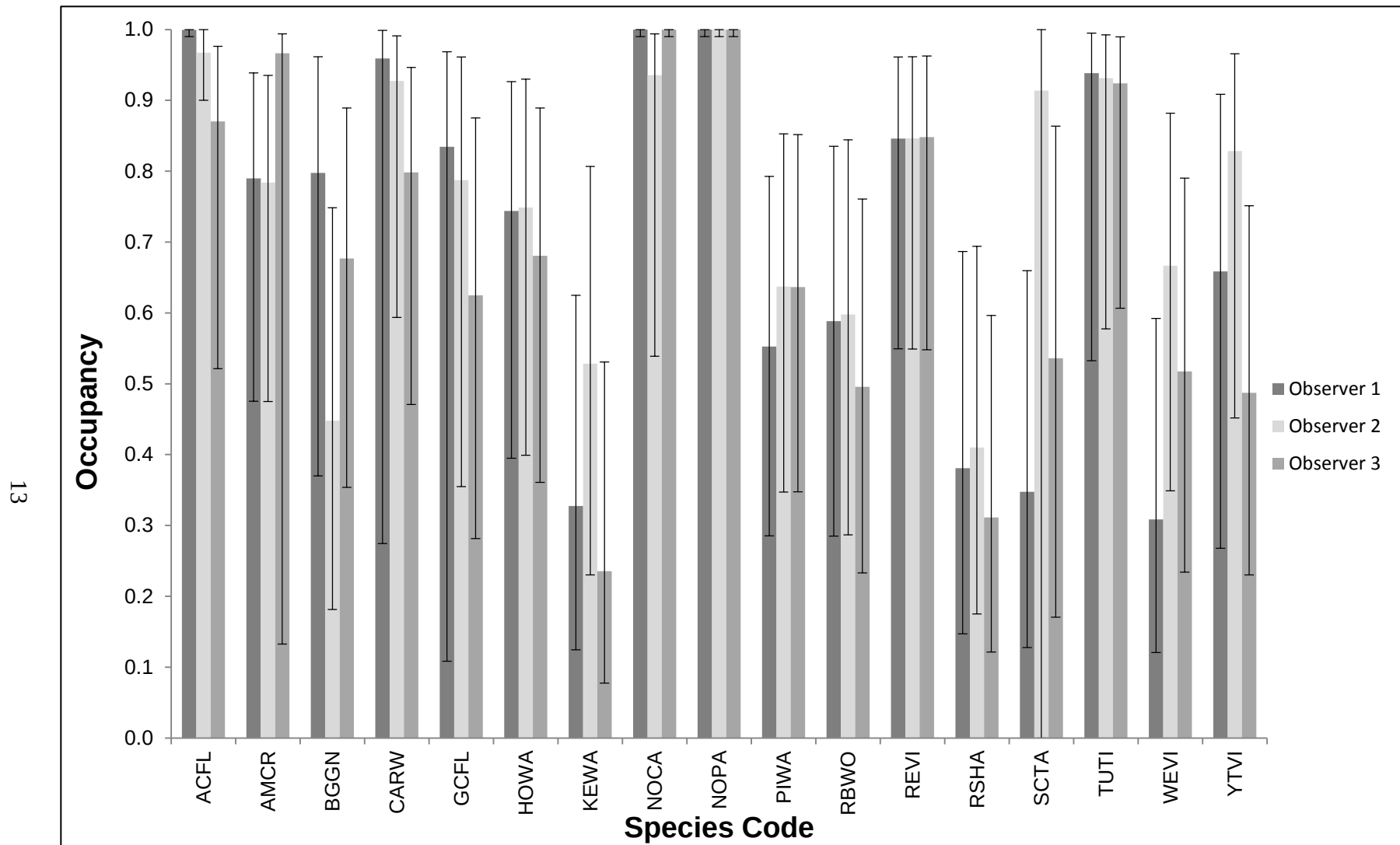


Figure 3. Occupancy estimates and 95% confidence intervals for species that met minimum sample size criteria for generating reliable estimates at HOBE (MacKenzie and Royle 2005). Results from three observers listening to identical recordings recorded at Horseshoe Bend National Military Park, 2012. Species codes adhere to four-letter codes from the American Ornithologists' Union (Appendix B).

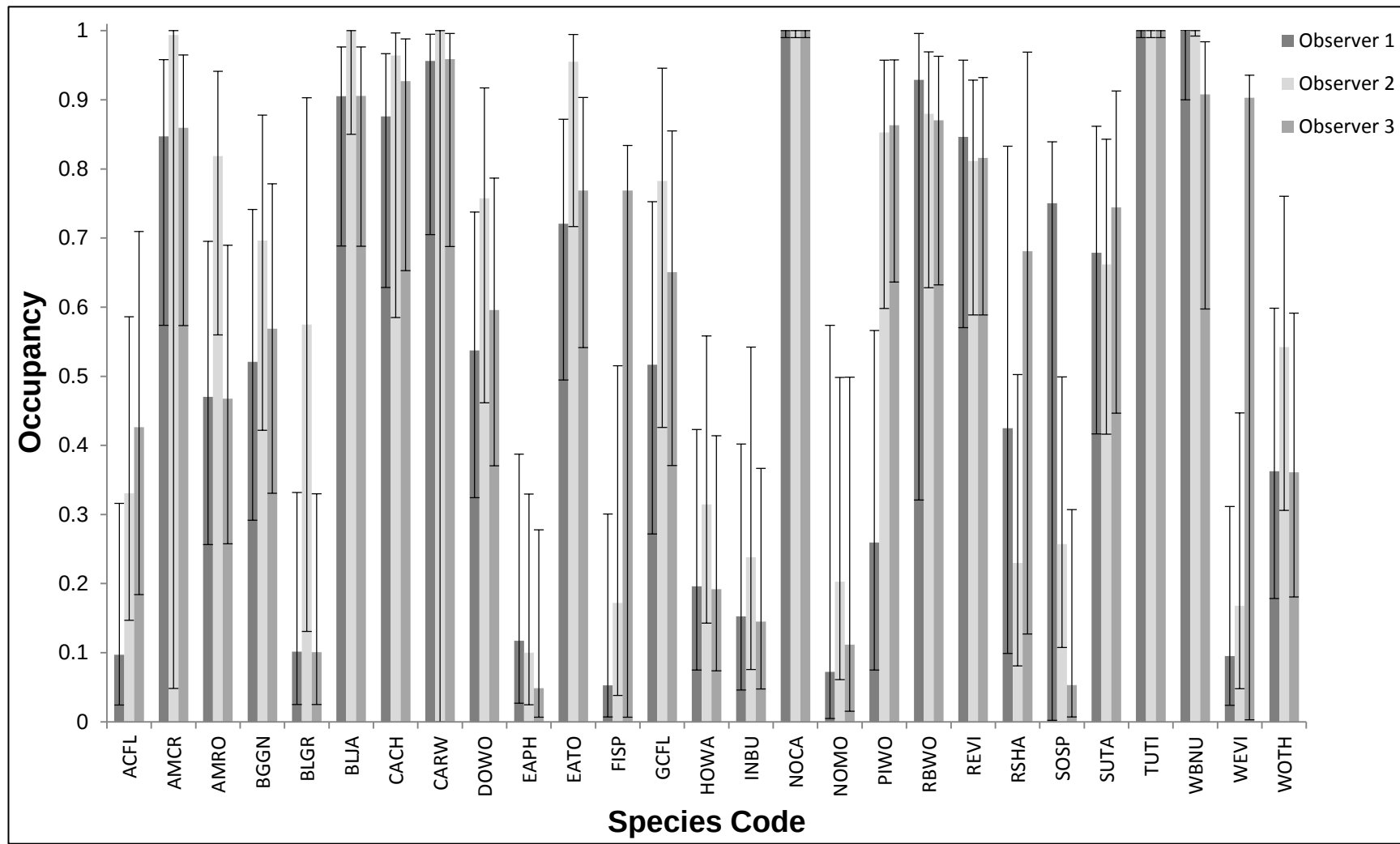


Figure 4. Occupancy estimates and 95% confidence intervals for species that met minimum sample size criteria for generating reliable estimates at KEMO (MacKenzie and Royle 2005). Results from three observers listening to identical recordings recorded at Kennesaw Mountain National Battlefield Park, 2012. Species codes adhere to four-letter codes from the American Ornithologists' Union (Appendix B).

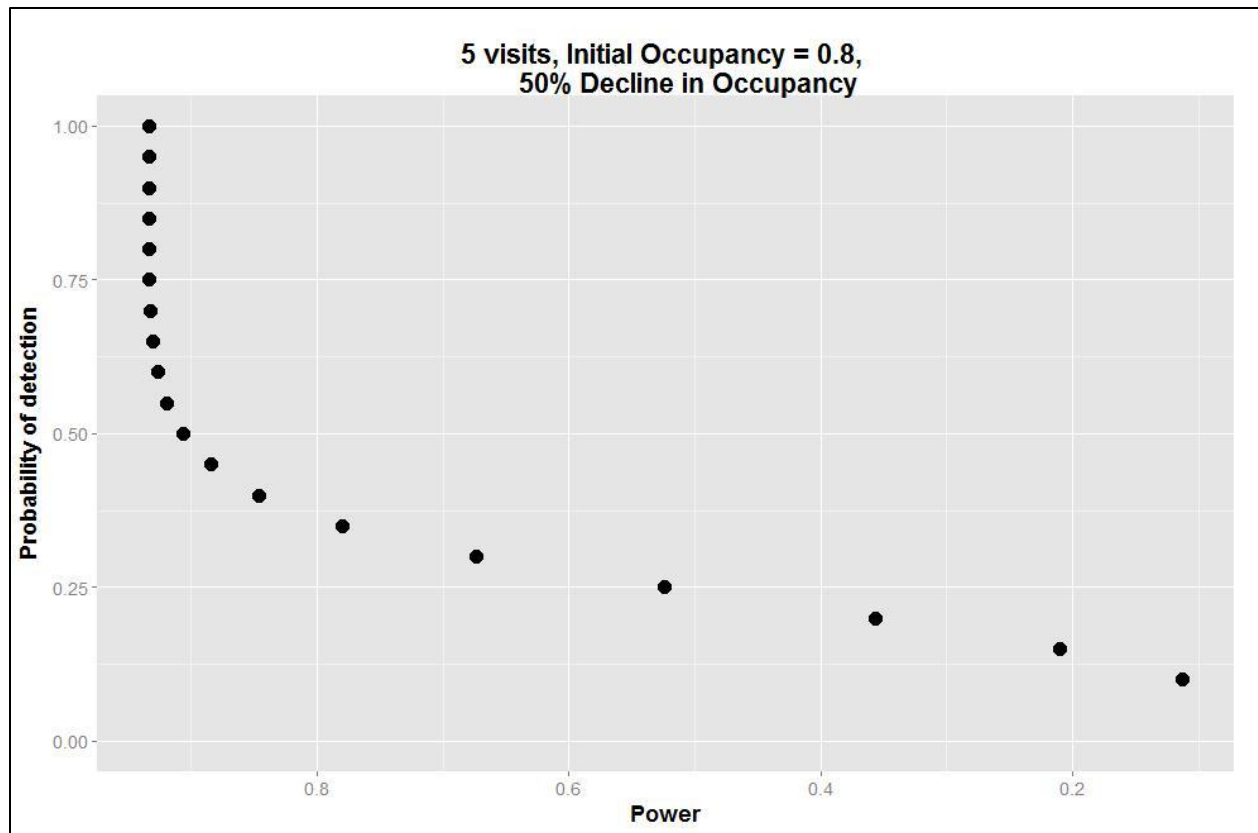


Figure 5. Power analysis for detecting a 50% decline in occupancy when initial occupancy is 0.8 and sites are visited five times (i.e., five recordings evaluated).

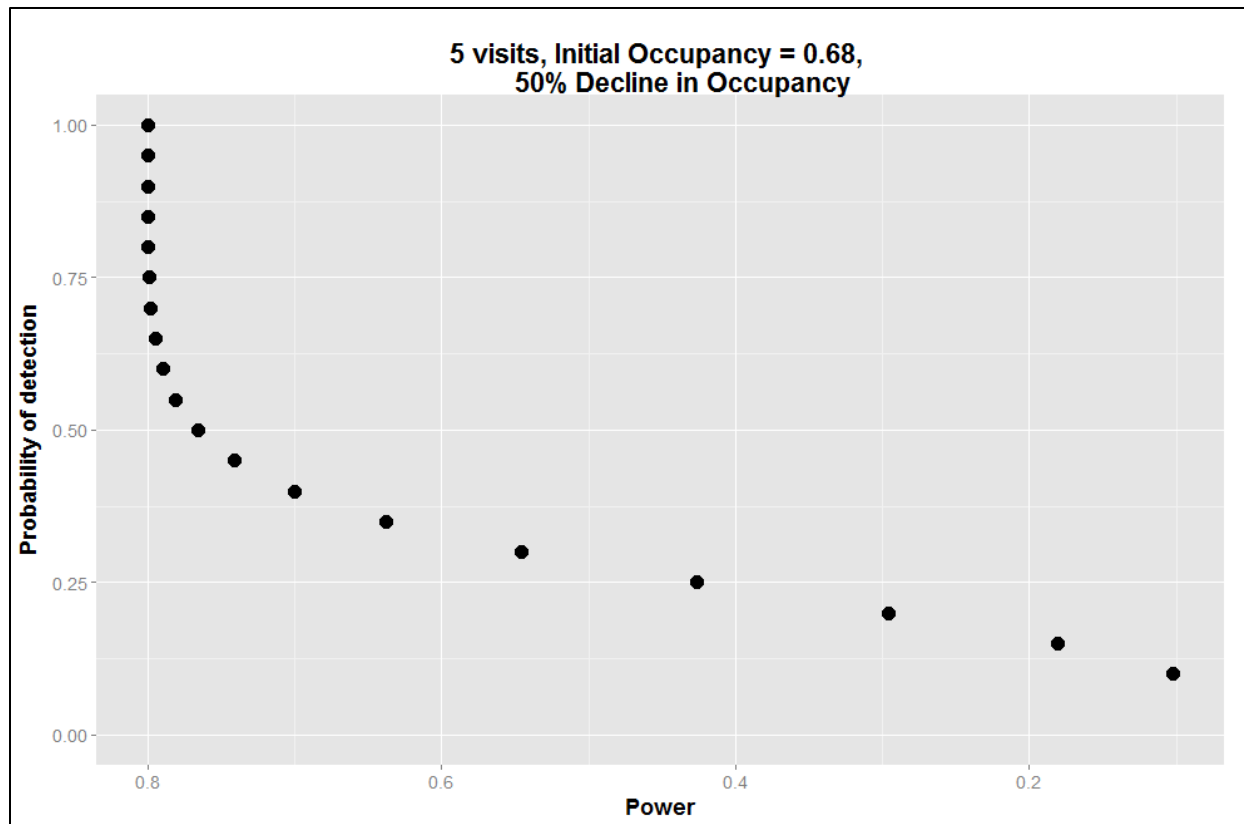


Figure 6. Power analysis for detecting a 50% decline in occupancy when initial occupancy is 0.68, and sites are visited five times (i.e., five recordings evaluated).

Conclusions and Recommendations

Despite variable levels of experience and expertise with bird vocalization identification, in this study we found moderate to high congruence across the three observers in identifying bird vocalizations. Mean differences across observers were lowest when estimates of occupancy and detection met the criteria for the number of visits as outlined by MacKenzie and Royle (2005); which are the most common species. Because these species are generally more common, observer accuracy and precision could possibly be higher because of a more frequent exposure to these species by all observers, both in the recordings and in everyday exposure outside of this study, and hence a superior familiarity with these species' vocalizations than those of less common species. Therefore, future trend analysis should focus on species that meet the minimum criteria to ensure the most reliable inferences.

The existing sampling and response designs are likely sufficient to characterize the majority of the bird community in terms of rates of occupancy. However, trend analysis for occupancy will be restricted to only a small percentage (i.e., < 10% in most cases—the most common species) of the bird community. Few species met the threshold for occupancy and probability of detection to achieve sufficient power to detect a decline of less than 50%. This highlights the importance of choosing specific species of management interest and targeting monitoring efforts to detect trends in those species. It is difficult, unrealistic, and often unattainable goal to monitor all species such that small to medium scale (i.e., < 50%) changes or trends are detected with minimal effort. Similar challenges often occur in the generation of a detection function for abundance and density estimation from point-count surveys with distance sampling where valid detection functions are only available for the most common species.

A possible remedy for low power is to pool detections across parks in order to increase the power to detect trends and produce more precise estimates. This requires making an assumption of similar detection between sites but may be a reasonable alternative given the small area and limited number of sampling locations of several parks in the Southeast Coast Network. Similar analysis approaches are currently used in other large-scale monitoring programs (IMBCR 2015).

Although detecting small differences or trends in individual species is the focus of some monitoring activities (e.g., species-specific conservation efforts and single-species management actions), it is not a priority of the NPS Southeast Coast Network Inventory and Monitoring Program.

Programmatically, rather, we are charged with determining the ecological condition of park resources, and because no single-species indicators of ecological condition exist in Southeast Coast Network park ecosystems, the protocol under which these data were collected was not designed or intended for single-species estimation (Byrne et al. 2014). Given this basic tenet, bird data are pooled based upon biologically relevant functional, compositional, and structural guilds, and are but one of the measures used to determine ecological condition of park resources. These bird data are evaluated in conjunction with co-located datasets that characterize the amphibian and vegetation communities.

While the Southeast Coast Network intends to develop algorithms or processes to facilitate automated identification of all species that occur on recordings (Byrne et al. 2014), this will likely

take several more years before they are tested and approved for monitoring. It is important to note, however, that once this point has been reached, an observer is still necessary to manually evaluate the putative recognitions. This will require a skilled observer to confirm or reject all recognitions that are identified by the automated recognition process. The results presented here indicate that well-trained observers should be able to produce results that are repeatable; which can only be furthered with the assistance of algorithms or processes that have been shown to work in similar bird-monitoring studies (e.g., Holmes et al. 2014).

Literature Cited

- Acevedo, M. A., and L. J. Villanueva-Rivera. 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* 34:211–214.
- Bardeli, R., D. Wolff, F. Kurth, M. Koch, K. H. Tauchert, and K. H. Frommolt. 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters* 31:1524–1534.
- Byrne, M. W., J. C. DeVivo, C. J. Wright, E. Thompson, and C. D. Jones. 2014. Protocol for monitoring landbird communities in Southeast Coast Network parks. Natural Resources Report NPS/SECN/NRR/—2014/853. National Park Service, Fort Collins, Colorado.
- Celis-Murillo, A., J. L. Deppe, and M. F. Allen. 2009. Using soundscape recordings to estimate bird species abundance, richness, and composition. *Journal of Field Ornithology* 80:64–78.
- Cunningham, R. B., D. B. Lindenmayer, H. A. Nix, and B. D. Lindenmayer. 1999. Quantifying observer heterogeneity in bird counts. *Australian Journal of Ecology* 24:270–277.
- DeVivo, J. C., C. J. Wright, M. W. Byrne, E. DiDonato, and T. Curtis. 2008. Vital signs monitoring in the Southeast Coast Inventory and Monitoring Network. Natural Resource Report NPS/SECN/NRR—2008/061. National Park Service, Fort Collins, Colorado.
- Diefenbach, D. R., D. W. Brauning, and J. A. Mattice. 2003. Variability in grassland bird counts related to observer differences and species detection rates. *The Auk* 120:1168–1179.
- Fiske, I., and R. Chandler. 2011. Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software* 43(10):1–23.
- Fitzpatrick, J. W., M. Lammertink, M. D. Luneau, T. W. Gallagher, B. R. Harrison, G. M. Sparling, K. V. Rosenberg, R. W. Rohrbaugh, E. C. Swarthout, and P. H. Wrege. 2005. Ivory-billed woodpecker (*Campephilus principalis*) persists in continental North America. *Science* 308:1460–1462.
- Goyette, J. L., R. W. Howe, A. T. Wolf, and W. D. Robinson. 2011. Detecting tropical nocturnal birds using automated audio recordings. *Journal of Field Ornithology* 82:279–287.
- Guillera-Arroita, G., and J. J. Lahoz-Monfort. 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution* 3(5):860–869.
- Haselmayer, J., and J. S. Quinn. 2000. A comparison of point counts and sound recording as bird survey methods in Amazonian southeast Peru. *The Condor* 102:887–893.

- Hobson, K. A., R. S. Rempel, H. Greenwood, B. Turnbull, and S. L. Van Wilgenburg. 2002. Acoustic surveys of birds using electronic recordings: new potential from an omnidirectional microphone system. *Wildlife Society Bulletin* 30:709–720.
- Holmes, S. B., K. A. McIlwrick, and L. A. Venier. 2014. Using automated sound recording and analysis to detect bird species-at-risk in southwestern Ontario woodlands. *Wildlife Society Bulletin* 38(3):591–598.
- Hutto, R. L., and R. J. Stutzman. 2009. Humans versus autonomous recording units: a comparison of point-count results. *Journal of Field Ornithology* 80:387–398.
- Integrated Monitoring in Bird Conservation Regions (IMBCR). 2015. Annual density and occupancy estimation with the RIMBCR package. Rocky Mountain Bird Observatory, unpublished report.
- Kendall, W. L., B. G. Peterjohn, and J. R. Sauer. 1996. First-time observer effects in the North American Breeding Bird Survey. *The Auk* 113:823–829.
- MacKenzie, D. I., and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology* 42:1105–1114.
- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, A. B. Knutson. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84:2200–2207.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probability rates are less than one. *Ecology* 83:2248–2255.
- McClintock, B. T., L. L. Bailey, K. H. Pollock, and T. R. Simons. 2010a. Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology* 91:2446–2454.
- McClintock, B. T., L. L. Bailey, K. H. Pollock, and T. R. Simons. 2010b. Experimental investigation of observation error in anuran call surveys. *Journal of Wildlife Management* 74:1882–1893.
- Miller, D. A., J. D. Nichols, B. T. McClintock, E. H. Campbell Grant, L. L. Bailey, and L. A. Weir. 2011. Improving occupancy estimation when two types of observational error occur: Nondetection and species misidentification. *Ecology* 92:1422–1428.
- NPSpecies. 2014. NPSpecies, The National Park Service Information on Species in National Parks. Secure online version. Available at: <https://irma.nps.gov/NPSpecies> (accessed June 2015).
- R Development Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://cran.r-project.org> (accessed June 2015).

- Rempel, R. S., J. M. Jackson, and J. N. Robinson. 2014. Acoustic monitoring and assessment of forest songbirds: Sample design, analysis methods, and observation error. CNFER Technical Report TR-012. Ontario Ministry of Natural Resources, Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario, Canada.
- Rempel, R. S., K. A. Hobson, G. Holborn, S. L. van Wilgenburg, and J. Elliott. 2005. Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital Processing Methods in the Laboratory. *Journal of Field Ornithology* 76:1–11.
- Royle, J. A., and W. A. Link. 2006. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* 87:835–841.
- Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *The Auk* 111:50–62.
- Theobald, D. M., D. L. Stevens, Jr., D. White, N. S. Urquhart, A. R. Olsen, and J. B. Norman. 2007. Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environmental Management* 40:134–146.
- U.S. Fish and Wildlife Service (USFWS). 2008. Birds of conservation concern 2008. U.S. Department of Interior, Fish and Wildlife Service, Division of Migratory Bird Management. Arlington, Virginia. Available at: <http://www.fws.gov/migratorybirds/> (accessed June 2015).
- Venier, L. A., S. B. Holmes, G. W. Holborn, K. A. McIlwrick, and G. Brown. 2012. Evaluation of an automated recording device for monitoring forest birds. *Wildlife Society Bulletin* 36:30–39.

Appendix A. Detection history discrepancy by park and observer

Table A-1. Detection discrepancy by species and observer (1 vs. 2) in CUIS 2012.

Park	Species	Observer 1	Observer 2
CUIS	AMRO	0	1
CUIS	BARS	0	1
CUIS	BBCU	0	1
CUIS	BWWA	0	1
CUIS	CAGO	0	1
CUIS	COGR	1	0
CUIS	FISP	0	1
CUIS	GBHE	1	0
CUIS	INBU	0	1
CUIS	OSPR	1	0
CUIS	ROYT	1	0
CUIS	RUBL	0	1
CUIS	SOSP	0	1
CUIS	WBNU	0	1
CUIS	WCSP	0	1
CUIS	YBCH	1	0

* "1" denotes species detected, "0" denotes species not detected

Table A-2. Detection discrepancy by species and observer (2 vs. 3) in CUIS 2012.

Park	Species	Observer 2	Observer 3
CUIS	AMGO	1	0
CUIS	BDOW	0	1
CUIS	AMRO	1	0
CUIS	BARS	1	0
CUIS	BBCU	1	0
CUIS	BWWA	1	0
CUIS	CAGO	1	0
CUIS	CAWA	0	1
CUIS	COGR	0	1
CUIS	CONI	1	0
CUIS	FICR	1	0
CUIS	FISP	1	0
CUIS	GRHE	1	0
CUIS	PUMA	1	0
CUIS	RUBL	1	0
CUIS	SOSP	1	0
CUIS	WCSP	1	0
CUIS	YBCH	0	1
CUIS	YBCU	1	0

* "1" denotes species detected, "0" denotes species not detected

Table A-3. Detection discrepancy by species and observer (1 vs. 3) in CUIS 2012.

Park	Species	Observer 1	Observer 3
CUIS	AMGO	1	0
CUIS	BDOW	0	1
CUIS	CAWA	0	1
CUIS	CONI	1	0
CUIS	FICR	1	0
CUIS	GBHE	1	0
CUIS	GRHE	1	0
CUIS	INBU	0	1
CUIS	OSPR	1	0
CUIS	PEAF	1	0
CUIS	PUMA	1	0
CUIS	ROYT	1	0
CUIS	WBNU	0	1
CUIS	YBCU	1	0

* "1" denotes species detected, "0" denotes species not detected

Table A-4. Detection discrepancy by species and observer (1 vs. 2) in HOBE 2012.

Park	Species	Observer 1	Observer 2
HOBE	AMRO	0	1
HOBE	CHSP	0	1
HOBE	CHSW	0	1
HOBE	EAPH	1	0
HOBE	FICR	0	1
HOBE	GBHE	1	0
HOBE	HAWO	1	0
HOBE	NOFL	0	1
HOBE	PUMA	1	0
HOBE	RHWO	1	0
HOBE	RTHA	1	0
HOBE	RTHU	1	0
HOBE	WBNU	1	0
HOBE	WITU	0	1
HOBE	YBCH	1	0

* "1" denotes species detected, "0" denotes species not detected

Table A-5. Detection discrepancy by species and observer (2 vs. 3) in HOBE 2012.

Park	Species	Observer 2	Observer 3
HOBE	AMRO	1	0
HOBE	BEKI	0	1
HOBE	CHSW	1	0
HOBE	FICR	1	0
HOBE	GBHE	0	1
HOBE	HAWO	0	1
HOBE	RHWO	0	1
HOBE	RTHA	0	1
HOBE	WITU	1	0
HOBE	YBCH	0	1

* "1" denotes species detected, "0" denotes species not detected

Table A-6. Detection discrepancy by species and observer (1 vs. 3) in HOBE 2012.

Park	Species	Observer 1	Observer 3
HOBE	BEKI	0	1
HOBE	CHSP	0	1
HOBE	EAPH	1	0
HOBE	NOFL	0	1
HOBE	PUMA	1	0
HOBE	RTHU	1	0
HOBE	WBNU	1	0

* "1" denotes species detected, "0" denotes species not detected

Table A-7. Detection discrepancy by species and observer (1 vs. 2) in KEMO 2012.

Park	Species	Observer 1	Observer 2
KEMO	BARS	1	0
KEMO	BHCO	1	0
KEMO	CEDW	1	0
KEMO	DOUB	1	0
KEMO	EAWP	0	1
KEMO	GRCA	1	0
KEMO	HAWO	1	0
KEMO	HOSP	0	1
KEMO	NOBO	0	1
KEMO	NOPA	0	1
KEMO	PUMA	1	0
KEMO	WCSP	0	1

* "1" denotes species detected, "0" denotes species not detected

Table A-8. Detection discrepancy by species and observer (2 vs. 3) in KEMO 2012.

Park	Species	Observer 2	Observer 3
KEMO	BEKI	0	1
KEMO	BHCO	0	1
KEMO	BHVI	1	0
KEMO	BRTH	1	0
KEMO	COGR	1	0
KEMO	COYE	1	0
KEMO	FICR	0	1
KEMO	GRHE	1	0
KEMO	HAWO	0	1
KEMO	HOSP	1	0
KEMO	LOWA	1	0
KEMO	NOBO	1	0
KEMO	RTHA	1	0
KEMO	RUBL	0	1
KEMO	WCSP	1	0
KEMO	WTSP	1	0

* "1" denotes species detected, "0" denotes species not detected

Table A-9. Detection discrepancy by species and observer (1 vs. 3) in KEMO 2012.

Park	Species	Observer 1	Observer 3
KEMO	BARS	1	0
KEMO	BEKI	0	1
KEMO	BHVI	1	0
KEMO	BRTH	1	0
KEMO	CEDW	1	0
KEMO	COGR	1	0
KEMO	COYE	1	0
KEMO	EAWP	0	1
KEMO	FICR	0	1
KEMO	GRCA	1	0
KEMO	GRHE	1	0
KEMO	LOWA	1	0
KEMO	NOPA	0	1
KEMO	PUMA	1	0
KEMO	RTHA	1	0
KEMO	RUBL	0	1
KEMO	WTSP	1	0

* "1" denotes species detected, "0" denotes species not detected

Appendix B. AOU species codes

Table B-1. Four-letter species codes as specified by the American Ornithologists' Union.

AOU Code	Common Name	AOU Code	Common Name
ACFL	Acadian Flycatcher	GRCA	Gray Catbird
AMCO	American Coot	GCFL	Great Crested Flycatcher
AMCR	American Crow	GRHE	Green Heron
AMGO	American Goldfinch	HAWO	Hairy Woodpecker
AMRE	American Redstart	HOWA	Hooded Warbler
AMRO	American Robin	HOFI	House Finch
BARS	Barn Swallow	HOWR	House Wren
BEKI	Belted Kingfisher	INBU	Indigo Bunting
BAWW	Black-and-white Warbler	KEWA	Kentucky Warbler
BTBW	Black-throated Blue Warbler	KILL	Killdeer
BLGR	Blue Grosbeak	LOSH	Loggerhead Shrike
BLJA	Blue Jay	LOWA	Louisiana Waterthrush
BGGN	Blue-gray Gnatcatcher	MAWA	Magnolia Warbler
BHVI	Blue-headed Vireo	MAWR	Marsh Wren
BTGR	Boat-tailed Grackle	MODO	Mourning Dove
BWHA	Broad-winged Hawk	NOBO	Northern Bobwhite
BRTH	Brown Thrasher	NOCA	Northern Cardinal
BHCO	Brown-headed Cowbird	NOFL	Northern Flicker
BHNU	Brown-headed Nuthatch	NOMO	Northern Mockingbird
CAGO	Canada Goose	NOPA	Northern Parula
CAWA	Canada Warbler	NRWS	Northern Rough-winged Swallow
CACH	Carolina Chickadee	OROR	Orchard Oriole
CARW	Carolina Wren	OSPR	Osprey
CEDW	Cedar Waxwing	OVEN	Ovenbird
CHSW	Chimney Swift	PABU	Painted Bunting
CHSP	Chipping Sparrow	PIWO	Pileated Woodpecker
CLSW	Cliff Swallow	PIWA	Pine Warbler
COGR	Common Grackle	PRAW	Prairie Warbler
CONI	Common Nighthawk	PROW	Prothonotary Warbler
COYE	Common Yellowthroat	PUMA	Purple Martin
COHA	Cooper's Hawk	RBWO	Red-bellied Woodpecker
DOWO	Downy Woodpecker	RBNU	Red-breasted Nuthatch
EABL	Eastern Bluebird	REVI	Red-eyed Vireo
EAKI	Eastern Kingbird	RHOW	Red-headed Woodpecker
EAME	Eastern Meadowlark	RSHA	Red-shouldered Hawk
EAPH	Eastern Phoebe	RTHA	Red-tailed Hawk
EATO	Eastern Towhee	RWBL	Red-winged Blackbird
EAWP	Eastern Wood-Pewee	RCKI	Ruby-crowned Kinglet
EUST	European Starling	RTHU	Ruby-throated Hummingbird
FISP	Field Sparrow	SCTA	Scarlet Tanager
FICR	Fish Crow	SOSP	Song Sparrow

Table B-1. Continued.

AOU Code	Common Name
SUTA	Summer Tanager
SWWA	Swainson's Warbler
TUTI	Tufted Titmouse
VEER	Veery
WBNU	White-breasted Nuthatch
WEVI	White-eyed Vireo
WTSP	White-throated Sparrow
WITU	Wild Turkey
WIFL	Willow Flycatcher
WODU	Wood Duck
WOTH	Wood Thrush
WOWA	Worm-eating Warbler
YWAR	Yellow Warbler
YBCU	Yellow-billed Cuckoo
YBCH	Yellow-breasted Chat
YRWA	Yellow-rumped Warbler
YTVI	Yellow-throated Vireo
YTWA	Yellow-throated Warbler

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS 910/128998, July 2015

National Park Service
U.S. Department of the Interior



Natural Resource Stewardship and Science
1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov

EXPERIENCE YOUR AMERICA TM