

Stewart, L. N., D. C. Tozer, J. M. McManus, L. E. Berrigan, and K. L. Drake. 2020. Integrating wetland bird point count data from humans and acoustic recorders. *Avian Conservation and Ecology* 15(2):9. <https://doi.org/10.5751/ACE-01661-150209>
Copyright © 2020 by the author(s). Published here under license by the Resilience Alliance.

Research Paper

Integrating wetland bird point count data from humans and acoustic recorders

Laura N. Stewart^{1,2}, Douglas C. Tozer², Janine M. McManus¹, Lucas E. Berrigan³ and Kiel L. Drake¹

¹Birds Canada, Saskatoon, SK Canada, ²Long Point Waterfowl and Wetlands Research Program, Birds Canada, Port Rowan, ON Canada, ³Birds Canada, Sackville, NB Canada

ABSTRACT. Wetland loss is cause for concern for populations of many wetland bird species throughout North America. The North American Breeding Bird Survey, the primary resource for broad-scale avian population data, does not provide sufficient data for many marsh bird species. Targeted marsh bird monitoring programs have been implemented across the continent in an attempt to fill this gap. Despite these efforts, a number of wetland species are so elusive that they remain an analytical challenge because of small sample sizes and low detectability. Thus, there is need for tools and approaches that will increase sampling efficiency and boost geographic representation. Autonomous recording units (ARUs) have the potential to address some of these challenges, but require the ability to combine in-person survey data with ARU data for collective analysis. Our primary objective was to estimate statistical offsets, or correction factors, to account for systematic differences between in-person and ARU counts of wetland-associated bird species. We found that ARU recordings were generally equivalent to in-person point counts, with bias in a small number of species (2 of 19 for Song MeterTM SM2 and 1 of 16 for Song MeterTM SM4 Acoustic Recorders; Wildlife Acoustics Inc. ©, Maynard, MA). However, bias was removed in all of the species through use of our correction factors. Therefore, our correction factors were effective for integrating in-person and ARU point count data even for species where differences exist. We also found that commercially available SM4 recorders have larger effective detection radii than SM2 recorders. Researchers should consider the microphone sensitivity and signal-to-noise ratios of any recording unit before purchasing, and more sensitive models with lower noise should be used where possible. Our results, and particularly our correction factors, are useful for biologists combining in-person and ARU point count data to achieve larger sample sizes, higher statistical power, and ultimately better information for more effective wetland conservation.

Intégration des données de comptage des oiseaux des zones humides obtenues par des enregistreurs humains et acoustiques

RÉSUMÉ. La disparition des zones humides est une cause de préoccupation pour nombreuses espèces d'oiseaux des zones humides à travers l'Amérique du Nord. La North American Breeding Bird Survey, principale source de données à grande échelle concernant les populations aviaires, s'avère toutefois insuffisante dans le cas de nombreuses espèces d'oiseaux des marais. Des programmes ciblés de surveillance des oiseaux des marais ont été mis en oeuvre sur l'ensemble du continent afin de combler cette lacune. En dépit de ces efforts, un certain nombre d'espèces d'oiseaux des marais sont si insaisissables qu'elles constituent un défi en termes d'analyse en raison de la petite taille des échantillons et de leur faible détectabilité. Il est donc nécessaire de trouver des outils et des approches qui augmenteront l'efficacité de l'échantillonnage et amélioreront la représentation géographique. Les unités d'enregistrement autonomes (ARU) pourraient permettre de relever certains de ces défis. Il faudrait toutefois être en mesure de combiner les données recueillies lors d'enquêtes en personne et les données ARU aux fins d'analyse collective. Notre objectif principal était d'estimer les décalages statistiques et d'appliquer des facteurs de correction afin de tenir compte de différences systématiques entre les comptages réalisés en personne et via des ARU sur les espèces d'oiseaux des zones humides. Nous avons constaté que les enregistrements des ARU étaient généralement équivalents aux comptages effectués en personne, avec un biais pour un petit nombre d'espèces (2 sur 19 avec les enregistreurs acoustiques Song MeterTM SM2 et 1 sur 16 avec les enregistreurs Song MeterTM SM4 ; Wildlife Acoustics Inc. ©, Maynard, MA). Nos facteurs de correction ont cependant permis d'éliminer ce biais pour toutes les espèces. En conséquence, nos facteurs de correction se sont avérés efficaces pour intégrer les données de comptage en personne et via des ARU, même pour les espèces où il existe des différences. Nous avons également constaté que les enregistreurs SM4 vendus dans le commerce présentent des rayons de détection effectifs plus importants que les enregistreurs SM2. Les chercheurs devraient tenir compte de la sensibilité des microphones et des ratios signal-bruit de toute unité d'enregistrement avant de faire un achat, et utiliser dans la mesure du possible les modèles les plus sensibles et les moins bruyants. Nos résultats et en particulier, nos facteurs de correction, sont utiles pour permettre aux biologistes qui associent des comptes en personne et via des ARU d'obtenir des échantillons de plus grande taille, une puissance statistique supérieure et finalement, de meilleures informations pour une conservation plus efficace des zones humides.

Key Words: ARU; autonomous recording unit; avian monitoring; bioacoustics; detection probability; detection radius; marsh bird; point count; survey bias

INTRODUCTION

Wetlands have declined in North America, in some areas dramatically, with emergent marsh suffering particularly high losses (Tiner 1984, Snell 1987, Dalh 2006). The primary causes of wetland loss are agriculture and development, which often also remove the natural upland habitats adjacent to marshes (Tiner 1984). Because of this habitat loss there is concern for populations of marsh birds such as bitterns, rails, and grebes that depend on emergent wetlands for breeding and stopover during migration (Saunders et al. 2019). Although population trends for many of these species are known in some regions (e.g., Tozer 2016), there are limitations with monitoring certain species throughout large parts of their range. The primary challenge is that many marsh birds have low detection rates, requiring several repeat visits to a large number of survey sites to obtain sufficient data for statistical modeling (Tozer et al. 2006, 2016, Steidl et al. 2013). Marsh birds are also elusive and crepuscular or nocturnal, typically necessitating night-time visits and/or call-broadcasts to entice individuals to call and be detected by observers (Tozer et al. 2006, Conway 2011). To further complicate monitoring efforts, many wetland complexes can be difficult or dangerous to access even in daylight, e.g., traversing floating mats of marsh vegetation on foot. Because of concerns about marsh bird populations and difficulties in monitoring these species, there is a need for tools and approaches that increase spatial and temporal efficiency and allow larger sample sizes.

Autonomous recording units (ARUs) have the potential to address some of these challenges by increasing detection of marsh birds and expanding the spatial and temporal capacity of monitoring programs (Shonfield and Bayne 2017). ARUs do have drawbacks, including high start-up cost, ongoing maintenance costs, potential equipment failure, and demands of data storage and management. However, many practitioners find the benefits of ARUs outweigh their disadvantages, and the use of ARUs is increasingly prevalent worldwide in the field of biodiversity monitoring (Haselmayer and Quinn 2000, Hobson et al. 2002, Steer 2010). ARUs have been shown to be effective tools for detecting marsh birds and for supplementing traditional monitoring programs (Klingbeil and Willig 2015, Drake et al. 2016, Znidarsic et al. 2020). ARUs can be deployed using the same spatial sampling scheme as in-person point counts, but eliminate the need for a human to be present at the time of observations. They can be programmed to make recordings of any specified duration at any time of the day or night, making many “revisits” possible at every sampling location. Furthermore, they can be deployed during winter when wetlands are frozen and easily accessible by anyone who is capable of navigating to a waypoint (Shonfield and Bayne 2017). Thus, ARUs overcome many of the wetland-specific challenges of traditional in-person surveys.

Research based on paired surveys shows that ARU surveys are generally equivalent to in-person point counts for species that are primarily detected using aural cues, although ARUs sometimes detect slightly fewer individuals and species on average (Alquezar and Machado 2015, Sedláček et al. 2015, Vold et al. 2017). Recent advances demonstrate that density estimates can be obtained from recordings by following the same standardization protocol that is used to integrate in-person point counts of varying lengths (Sólymos et al. 2013, Van Wilgenburg et al. 2017, Bombaci and Pejchar 2019). This protocol accounts for imperfect detection for

both in-person and ARU surveys, and applies a species-specific correction factor to data from ARU surveys. The correction factor describes the difference between the effective detection radii of human observers and ARU recordings. Given the potential of ARUs to enhance marsh bird monitoring efforts, it is important to quantify the correction factors so that bird survey data from either source can be combined for analysis.

Here we apply the framework for paired point counts created by Van Wilgenburg et al. (2017) to a suite of birds inhabiting marshes and their adjacent uplands. In doing so, our primary objective is to estimate statistical offsets, or correction factors, to account for systematic differences between in-person and recording-based counts of wetland-associated bird species. As a secondary objective, we also test for differences in correction factors between two commercial recorders in widespread use, the Song Meter™ SM2 and Song Meter™ SM4 Acoustic Recorders (Wildlife Acoustics Inc. ©, Maynard, MA). As bioacoustic technology advances, the internal noise of recording units is decreasing and the microphone sensitivity is increasing (Wildlife Acoustics 2011, 2019). Therefore, we anticipate correction factors to vary among models of ARU as well as among bird species that vocalize at various frequencies. We predicted that mean counts from ARU surveys would be slightly less than mean counts from in-person surveys, and that SM2 counts would be slightly less than SM4 counts because of differences in their microphone sensitivities and respective effective detection radii. Results from this work and associated correction factors will be useful for researchers who wish to combine wetland bird survey monitoring data from in-person and ARU sources for combined analysis.

METHODS

Sampling

We conducted this study in the Canadian provinces of Saskatchewan, Ontario, Prince Edward Island, New Brunswick, and Nova Scotia in 2018 and 2019. All surveys were completed during weather favorable for detecting birds (no precipitation, wind < 20 km/h) at wetlands dominated by emergent vegetation, i.e., marshes, during the marsh bird breeding season from 23 May to 29 June. In total 602 surveys were conducted by 13 observers at 480 survey sites across 94 wetlands or wetland complexes. Of these 480 sites, 301 were part of existing monitoring programs and 179 were selected opportunistically. Most sites were located on roads directly adjacent to wetlands (~90%), some were off-road at the edge of a wetland and its adjacent upland (~5%), and a small minority were in the middle of large marsh complexes (~5%). Upland vegetation included prairie, agriculture, parkland, and forest.

Each in-person survey was a passive point count, i.e., no call-broadcast was used, lasting five minutes. Marsh bird surveys typically include a call-broadcast sequence (Conway 2011) following the passive segment to evoke calls from birds that otherwise might remain undetected. However, the use of call-broadcasts would have influenced the movement and singing rate of individual birds and thus violated the assumptions of distance and time removal sampling, so we compared only results from passive surveys. In the field, we recorded observations of a group of focal species. The minute interval (i.e., minute 0, 1, 2, 3, or 4) and the distance band (0–50 m, 50–100 m, or > 100 m) was noted

for each individual at the time the bird was first detected. Focal species were Pied-billed Grebe (*Podilymbus podiceps*), Red-necked Grebe (*Podiceps grisegena*), Virginia Rail (*Rallus limicola*), Sora (*Porzana carolina*), Common Gallinule (*Gallinula galeata*), American Coot (*Fulica americana*), Wilson's Snipe (*Gallinago delicata*), American Bittern (*Botaurus lentiginosus*), Least Bittern (*Ixobrychus exilis*), Sedge Wren (*Cistothorus platensis*), Marsh Wren (*Cistothorus palustris*), Clay-colored Sparrow (*Spizella pallida*), LeConte's Sparrow (*Ammodramus leconteii*), Nelson's Sparrow (*Ammodramus nelsoni*), Savannah Sparrow (*Passerculus sandwichensis*), Song Sparrow (*Melospiza melodia*), Swamp Sparrow (*Melospiza georgiana*), Yellow-headed Blackbird (*Xanthocephalus xanthocephalus*), Red-winged Blackbird (*Agelaius phoeniceus*), Brown-headed Cowbird (*Molothrus ater*), Common Yellowthroat (*Geothlypis trichas*), and Yellow Warbler (*Setophaga petechia*). Environmental information was taken at each survey station including wind speed and noise code (0 = none, 1 = light, 2 = moderate, 3 = heavy, 4 = unusable).

Observers recorded the full length of each in-person survey in the field using an SM2 recorder with a pair of SMX-II microphones, and surveyors in Saskatchewan used an SM4 in addition to the SM2. Sensitivity of SM2 microphones was tested using an Extech Model 407744 Sound Level Calibrator and those that tested below -42 dB were not used in the study (Turgeon et al. 2017). SM4 microphones were tested using the same instrument, and all tested above -35 dB. ARUs were affixed to a tripod at a height of approximately 1.3 m during recordings of surveys. ARUs were programmed to record in stereo WAV file format, with a sampling rate of 44100 Hz and factory default settings for the microphone preamplifier. Observers stood 3 to 5 m away from the ARUs to minimize recording extraneous noise.

To minimize observer bias, we transcribed ARU recordings at the end of the field season with the same observers interpreting the recordings as who had conducted the paired in-person point counts. Observers did not consult field notes, and we randomized file names so that information about date and location was not known to the observer during transcription. During transcription, we noted individuals of focal species and the minute interval of first detection. Transcription was done using software that allowed observers to watch spectrograms as they listened to recordings (e.g., Audacity®, Raven Pro (© Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Ithaca, NY). Unlike under field conditions, transcribers were allowed to pause and rewind recordings to confirm identification, as well as consult with colleagues for difficult identifications (which occurred less than 10 times during our study), as is typically done in audio transcription. Although the differences in methodology between in-person and ARU surveys did make the comparison between humans and ARUs slightly less equal, it made our correction factors more accurate for realistic data collection. A noise code was assigned to each recording on the same five-point scale as in the field and paired-surveys that had recordings with excessive noise interference were not included in our analysis.

Framework

Sólymos et al. (2013) demonstrated that data from point counts of varying lengths and radii can be adjusted using time-of-detection (time removal) and distance sampling so that they can be combined into larger datasets without creating systematic

biases. The expected count at a point count location is a function of a species' availability (p) and perceptibility (q). Availability is the probability that an individual gives a cue during the survey period, given it is present, and is a function of its singing rate (ϕ). Time removal models can be fitted to estimate singing rates, and therefore availability (Farnsworth et al. 2002, Sólymos et al. 2013). These models fit time to first detection (t_{ij}) to an exponential function that is described by ϕ , assuming a constant singing rate:

$$f(t_{ij}) = \phi e^{-t_{ij}\phi} \quad (1)$$

Thus p can be calculated by the cumulative density function from time $t = 0$ to $t = t_{ij}$:

$$p(t_{ij}) = 1 - e^{-\phi t_{ij}} \quad (2)$$

Where t_{ij} is the duration, in minutes, of the point count. Perceptibility (q) is the probability that an individual is detected, given it is available, and depends on its effective detection radius (τ). Distance sampling models can be fitted to estimate effective detection radii, and therefore perceptibility (Buckland et al. 2001, Matsuoka et al. 2012). These models fit distance of individual birds (r) to a half-normal distribution that is described by τ :

$$g(r) = e^{-\frac{r^2}{\tau^2}} \quad (3)$$

We assumed that all available birds at $r = 0$ were detected, that r was measured without error, and that birds were detected at their initial location.

The paired point count method of Van Wilgenburg et al. (2017) assumes that true bird abundance (N) is equal for in-person and ARU surveys. It models differences in detection by making the simplifying assumption that in-person and ARU surveys have equal availability ($p_H = p_A$), but allows for differences in effective detection radii ($\tau_H \neq \tau_A$). The assumption that $p_H = p_A$ is then explicitly tested using time removal models. We did not model for differences in p , but we noted that this was possible and may be necessary for some species. We also reasoned that there were at most only very small differences in false positive or negative misidentifications during in-person compared to ARU surveys, so we did not adjust for such differences (Rempel et al. 2019). For complete methodology, see Van Wilgenburg et al. (2017). Briefly, the expected count for a species for an in-person (Y_H) survey at any point count location was:

$$E[Y_H] = Npq \quad (4)$$

Where N is the species' true abundance (number of birds), p is availability, and q is perceptibility. We note that N is not observed directly. This can be rewritten as:

$$E[Y_H] = D_H A_H p \quad (5)$$

Where D is point-level density (birds/area) and A_H is the area sampled by a human observer. The exact area sampled is not known but the effective area sampled can be estimated using distance sampling. The in-person effective area sampled is described by:

$$A_H = \pi \tau_H^2 \quad (6)$$

where τ_H is the in-person effective detection radius. We assumed perfect detectability ($q = 1$) within this distance.

We were interested in the mean ratio of in-person (Y_H) to ARU (Y_A) expected survey counts:

$$\frac{E[Y_A]}{E[Y_H]} = \frac{D_A A_A p_A}{D_H A_H p_H} = \frac{D_A \pi \tau_A^2 p_A}{D_H \pi \tau_H^2 p_H} \quad (7)$$

We assumed density and availability were equal between survey methods ($D_A = D_H$ and $p_A = p_H$), and we defined a new variable:

$$\delta = \frac{\tau_A}{\tau_H} \quad (8)$$

then

$$\frac{E[Y_A]}{E[Y_H]} = \frac{\tau_A^2}{\tau_H^2} = \left(\frac{\tau_A}{\tau_H}\right)^2 = \delta^2 \quad (9)$$

Where δ^2 is the squared correction factor that describes the difference between ARU and in-person detection radii. This can be estimated as above, by comparing the mean counts of ARUs and in-person surveys, but it can also be estimated by back-transforming the fixed-effect “survey type” coefficient of Poisson or negative binomial generalized linear (GLM) or generalized linear mixed (GLMM) models (i.e., $\delta^2 = \exp[\beta]$).

Statistical models of count data can be fit using statistical offsets, or estimated correction factors (\hat{C}_i), that transform total counts to bird densities as individuals per unit area, taking imperfect detection into account. We defined the correction factor as:

$$\hat{C} = \pi \hat{\tau}_H^2 \hat{p} \quad (10)$$

For Poisson GLMs, the offset will be log-transformed (offset = $\log[\hat{C}_i]$). A mean count at any given in-person survey location ($\lambda_{i,H}$) can be expressed as:

$$\lambda_{i,H} = D_i \pi \hat{\tau}_H^2 \hat{p} = D_i \hat{C}_{i,H} \quad (11)$$

The calculated δ^2 can then be incorporated into statistical offsets for ARU surveys and thus eliminate systematic biases in calculated bird densities. GLMM models can incorporate both human and ARU surveys when they are fit with these correction factors and include the δ^2 coefficient and I_A , an indicator function taking on a value of 0 for in-person surveys and 1 for ARU surveys:

$$\log(\lambda_i) = \log(D_i) + \log(\hat{C}_{i,H}) + I_A \log(\delta^2) \quad (12)$$

Data analysis

Data analysis followed Van Wilgenburg et al. (2017). Species with fewer than 14 detections on either in-person or ARU surveys were excluded from analysis. We began by estimating $\hat{\tau}_H$ and singing rates for each species by fitting distance and removal models using the “detect” R package (Sólymos et al. 2018) as described by Sólymos et al. (2013). Time removal models were fitted separately for ARU surveys ($\hat{\phi}_A$), in-person surveys ($\hat{\phi}_H$), and the full dataset using survey type as factor ($\hat{\phi}$). Analysis was not continued for species where the full dataset models had survey type factors whose 95% confidence intervals overlapped 0. We used the calculated τ_H and $\hat{\phi}$ to estimate in-person correction factors (\hat{C}_H).

We estimated the correction factor (δ) and validated models by using repeated random subsampling of the data. For each species,

we randomly selected data from 70% of survey stations where the species was present to develop GLMs to estimate δ . The remaining 30% of data were used to calculate δ again by dividing the mean count from ARU surveys by the mean count from in-person surveys. This was repeated 50 times and we calculated means and 95% confidence intervals for both measures of δ across all replicates. We examined the 95% confidence intervals and correlation of the GLM-based calculations of δ to the empirical ratios to validate the performance of the models.

We also used the 30% validation set to examine whether the inclusion of δ in statistical offsets reduced bias in predicted densities from in-person versus ARU surveys. We first fit a GLM with a statistical offset $\log(\hat{\tau}_H^2 \cdot \pi \cdot \hat{p})$, i.e., $\log(\hat{C}_H)$, to calculate average density using in-person data. We then fit two competing models using ARU data, one with the same statistical offset and the other including δ : $\log([\hat{\delta} \cdot \hat{\tau}_H]^2 \cdot \pi \cdot \hat{p})$. We calculated bias as the difference in estimated density between the in-person model and each competing ARU model. We used GLMs instead of GLMMs like Van Wilgenburg et al. (2017) because the inclusion of a random effect at the site level did not improve the models nor change our results.

We repeated the above analysis twice, once for paired in-person and SM2 surveys and once for paired in-person and SM4 surveys. Thus, estimates of δ were calculated separately for SM2 surveys and SM4 surveys.

We used the full dataset to create competing Poisson generalized linear mixed models (GLMMs) to predict species abundance using site as a random effect. We included a null model (intercept only), a model with fixed effect for survey type (three levels: in-person, SM2, and SM4), a model with noise as a fixed effect, a model with both noise and survey type as a fixed effect (survey type + noise), and a model that incorporated the interaction between noise and survey type (survey type * noise). Noise level (0–4) was treated as a linear covariate. We did not test for a year effect because we compared counts conducted at the exact same time and place; therefore, year effect is controlled for by the paired design. We used Akaike’s Information Criterion (AIC) to select the most parsimonious of the five models (Akaike 1973, Beier et al. 2001). Next, we calculated δ by taking the median, 2.5% and 97.5% quantiles from 10,000 Monte Carlo simulations drawn from a multivariate normal distribution using survey type GLMM coefficients as the means and variance-covariance matrices. Last, we plotted noise as a function of wind speed (km/h) for each of the three survey types separately, and calculated linear lines of best fit to explore how noise varied in accordance to wind speed depending on the type of survey.

RESULTS

Two focal species did not meet the criteria for minimum number of detections for SM2 paired surveys (Common Gallinule and Least Bittern) and four did not meet criteria for SM4 paired surveys (Common Gallinule, Least Bittern, Nelson’s Sparrow, and Virginia Rail; Table A1.1). Effective detection radii for in-person counts ranged from 51 m to 181 m (Fig. 1). Availability across all species ranged from 0.171 to 1, and correlation in availability between in-person and ARU counts was weak but significant (SM2: Pearson’s $r = 0.58$, $p = 0.007$; SM4: Pearson’s $r = 0.65$, $p = 0.003$; Fig. 2). One species did not meet the assumption of equal availability for SM2 surveys (Song Sparrow, survey effect for ARU

$\beta = 0.49$, SE = 0.18), and two did not meet the assumption for SM4 surveys (Brown-headed Cowbird, model failed to converge, and Song Sparrow, survey effect for ARU $\beta = 0.65$, SE = 0.19).

Fig. 1. Effective detection radii calculated using in-person data (τ_H). Means and 95% confidence intervals are shown. Estimates of τ_H were calculated using distance sampling models that fit data to a half-normal detection function. Species are sorted taxonomically. See main text for scientific species names.

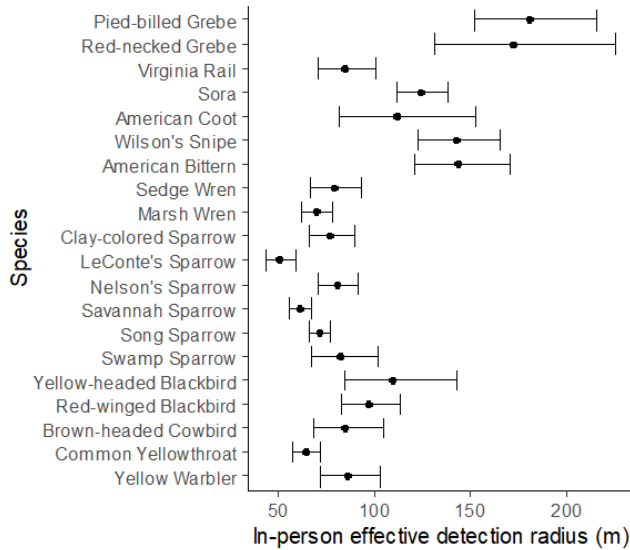
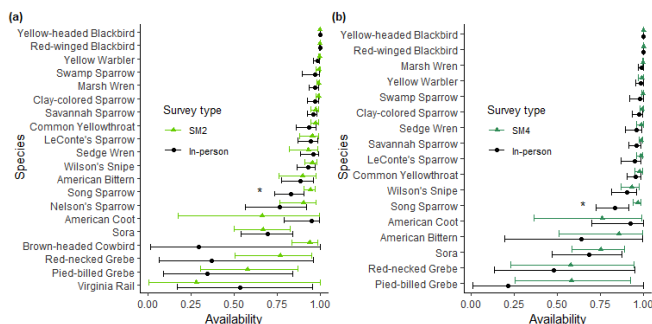


Fig. 2. Availability (p) estimates, calculated as the probability that an individual will give a cue during a 5-minute point given that it is present. Availability was calculated separately for in-person and autonomous recording unit (ARU) surveys. Means and 95% confidence intervals are shown in (a), SM2, and (b), SM4. Asterisks indicate species where in-person availability was significantly different from ARU availability. See main text for scientific species names.



The median estimate of δ across all species for SM2 surveys was 0.98 and the median estimate of δ for SM4 surveys was 1.03. Confidence intervals of δ for SM2 overlapped 1 for 13 of 19 species, and δ for SM4 overlapped 1 for 13 of 16 species (Fig. 3; Tables A1.5 and A1.6). SM4 δ estimates were greater than SM2

δ estimates for 14 of 16 species. Across species, estimates of δ using calibration data were highly correlated with estimates of δ using validation data (Fig. 4; Pearson's $r = 0.99$, $p < 0.001$ for both SM2 and SM4). Within each species, estimates of species density using in-person data for each of the 50 repeated random subsamples were well correlated with estimates of density using ARU data, with the exception of a few species: American Coot, Pied-billed Grebe, Red-winged Blackbird, and Yellow-headed Blackbird (Fig. A1.1).

Fig. 3. Correction factors (δ) for SM2 and SM4 surveys. Where δ is less than 1, autonomous recording units (ARUs) count fewer individuals on average than in-person surveys. Presented are medians and 95% confidence intervals of Monte Carlo simulations (see Methods). See main text for scientific species names.

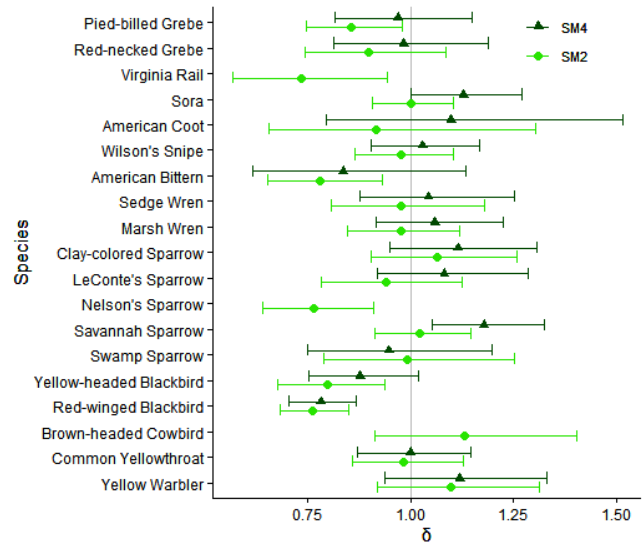
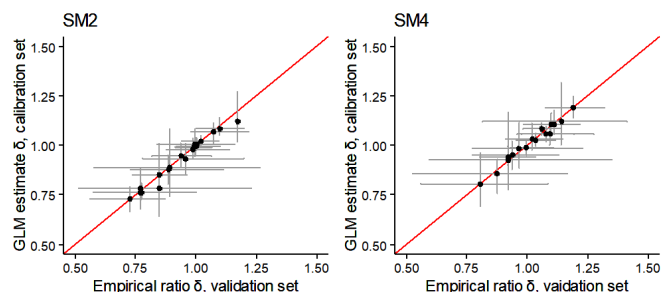
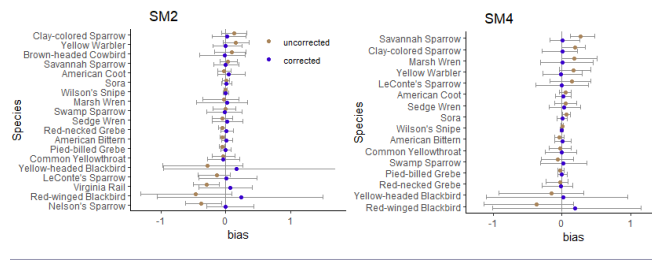


Fig. 4. Comparison of calibration estimates of δ (Poisson generalized linear model [GLM] fixed-effect estimates) and validation estimates of δ (empirical ratios of mean counts). Estimations were made by taking 50 repeated random subsamples of the data, fitting GLMs to 70% of the data, and calculating empirical ratios with the remaining 30%. Comparisons are shown for both SM2 (left) and SM4 (right). The red line indicates 1:1 correspondence.



Including δ estimates into statistical offsets reduced bias for all 19 species for SM2s and all 16 species for SM4s (Fig. 5). Of the 10 species with negative biases for SM2 surveys, two had 95% confidence intervals that did not overlap 0 (Virginia Rail and Nelson's Sparrow). For SM4 surveys, no birds had a significant negative bias, and of eight species with positive biases, one had a 95% confidence interval that did not overlap 0 (Savannah Sparrow).

Fig. 5. Bias in estimated densities from autonomous recording unit (ARU) surveys (top, SM2; bottom, SM4) compared to bias from the same surveys but including the correction factor δ as statistical offset. Bias was calculated as difference in predicted bird density (birds / ha) from in-person point counts to ARU point counts. In-person and uncorrected density was calculated using the QPAD approach (Sólymos et al. 2013) of applying species-specific offsets ($\log[\tau_H^2 \cdot \pi \cdot p]$) to raw counts of species abundance. Corrected density was calculated using the same offset but including the correction factor δ : $\log[\delta \cdot \tau_H^2 \cdot \pi \cdot \varphi]$. Biases were calculated over 50 repeated random subsamples, using 70% of data to calculate δ estimates for use in corrected offsets and using the remaining 30% to calculate estimated in-person, corrected ARU, and uncorrected ARU densities. See main text for scientific species names.



Based on AIC model selection, the noise model was most parsimonious for 11 species, the null model most parsimonious for 4 species, the survey type + noise model most parsimonious for 4 species, the survey type model most parsimonious for 1 species, and the survey type-noise interaction model received no support (Table 1). There was no effect of wind on noise for in-person surveys ($\beta = -0.035$, $SE = 0.021$), but there was a significant positive effect for SM2 surveys ($\beta = 0.165$, $SE = 0.028$) and SM4 surveys ($\beta = 0.108$, $SE = 0.035$).

DISCUSSION

Our results confirm the growing consensus that ARU recordings are generally equivalent to in-person point counts, which is shown by other studies (Alquezar and Machado 2015, Sedláček et al. 2015, Vold et al. 2017) and meta-analyses (Shonfield and Bayne 2017, Darras et al. 2018). Some studies show that in-person point counts detect more birds than ARU recordings (e.g., Borker et al. 2015, Sidie-Slettedahl et al. 2015), but most of these studies used older technology, e.g., SM1, with less sensitive microphones than newer models (Rempel et al. 2013), or they transcribed recordings using automated recognizers rather than manually. These studies may, therefore, reflect limitations of technology or methodology at the time of publication rather than the utility of ARUs in general. In contrast, studies comparing a typical sampling of an ARU deployment, i.e., manual transcription of multiple

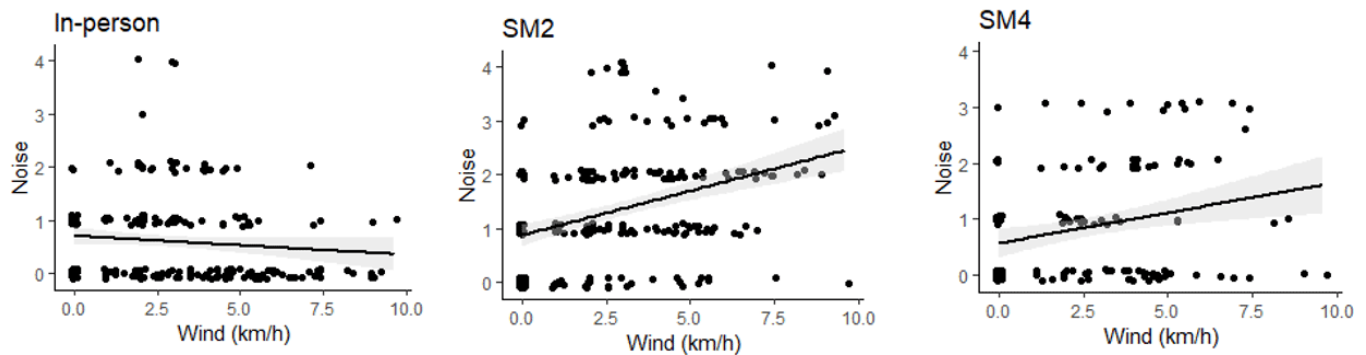
recordings per station, to a typical in-person point count (commonly one visit per station) generally found that ARUs perform better than in-person surveys, especially for nocturnal birds (Zwart et al. 2014, Klingbeil and Willig 2015, Bobay et al. 2018). This is not surprising, considering the greater number and broader diurnal distribution of sampling occasions possible from ARU recordings relative to single in-person counts. This is something we were able to control for with our paired human-ARU point counts, which had equal amounts of sampling time. Our findings also show that the newer SM4s are more sensitive and have larger effective detection radii than SM2s, a finding that is consistent with previous research showing that newer recorder technology improves on older models (Rempel et al. 2013, Yip et al. 2017). Therefore, the make and model of recording unit needs to be accounted for by practitioners using ARU in monitoring programs.

Table 1. Difference in Akaike's Information Criterion (ΔAIC) values for Poisson generalized linear mixed models (GLMMs) predicting species abundance that rank models relative to the most parsimonious. ΔAIC values of 0 represent the most parsimonious models. Null was an intercept-only model. Survey type consisted of three levels: in-person, SM2, and SM4. See main text for scientific species names.

Species	Null	Survey type	Noise	Survey type + Noise	Survey type * Noise
Pied-billed Grebe	40.91	40.40	0.00	2.66	5.45
Red-necked Grebe	15.51	18.15	0.00	3.37	5.53
Virginia Rail	44.00	41.22	5.00	0.00	3.79
Sora	91.29	90.86	0.00	0.19	3.28
American Coot	4.96	7.93	0.00	2.03	5.16
Wilson's Snipe	46.50	49.90	0.00	2.97	4.29
American Bittern	55.35	55.06	0.00	1.90	5.27
Sedge Wren	16.68	20.18	0.00	1.75	5.01
Marsh Wren	78.64	81.47	0.00	2.69	5.57
Clay-colored Sparrow	0.00	2.19	1.96	4.01	7.64
LeConte's Sparrow	9.17	10.77	0.00	1.67	5.44
Nelson's Sparrow	55.55	49.90	0.24	0.00	3.08
Savannah Sparrow	42.32	37.85	3.29	0.00	0.80
Song Sparrow	0.00	3.97	1.64	5.57	9.19
Swamp Sparrow	6.84	10.59	0.00	3.94	6.63
Yellow-headed Blackbird	22.28	18.40	14.83	0.00	3.93
Blackbird					
Red-winged Blackbird	28.32	0.00	25.61	1.84	0.99
Blackbird					
Brown-headed Cowbird	0.00	2.72	1.97	4.64	8.56
Cowbird					
Common Yellowthroat	35.93	39.86	0.00	3.45	7.21
Yellow Warbler	0.00	2.13	1.90	2.61	5.59

Density estimates produced from ARU recordings were generally equivalent to those produced by in-person surveys (Fig. A1.1), and most estimates of δ had 95% confidence intervals that overlapped 1 (Fig. 3). In-person and uncorrected ARU density estimates from repeated random subsampling were highly correlated for most species, and for many species were nearly equivalent (Fig. A1.1). Models for most species did not include an ARU effect, and bias for most species overlapped one, further suggesting that in-person and ARU counts are equivalent

Fig. 6. Wind and noise level for in-person, SM2, and SM4 surveys. Points are jittered to reduce overplotting and linear models are shown.



for most species (Fig. 3; Table 1). However, when present, both minor and major biases were corrected by inclusion of δ , demonstrating that these correction factors will facilitate effective integration of in-person and ARU surveys.

Species with negative biases and δ values less than 1 were: for SM2, Pied-billed Grebe, Virginia Rail, American Bittern, Nelson's Sparrow, Yellow-headed Blackbird, and Red-winged Blackbird; and for SM4, Yellow-headed Blackbird and Red-winged Blackbird (Figs. 3 and 5). Both blackbird species were detected at high densities, and an accurate count without visual cues became impossible beyond a certain threshold (see Table A1.2 and Fig. A1.1). Specifically, recordings did not estimate densities higher than 1.0 birds / ha, and we never estimated more than 10 individual blackbirds of either species from a recording. By contrast, in-person surveys detected as many as 30 Yellow-headed Blackbirds and as many as 50 Red-winged Blackbirds. The limitations of recordings for estimating high densities is also shown by Drake et al. (2016), who found that transcribers underestimated the number of Yellow Rails (*Coturnicops noveboracensis*) when there were more than seven individuals vocalizing on a single recording. Furthermore, while the large majority (> 97%) of detections of most secretive marsh birds are aural, detections of a few marsh birds such as Horned Grebe (*Podiceps auritus*) and Eared Grebe (*Podiceps nigricollis*) are over 90% visual (K. L. Drake, unpublished data). Thus, in-person surveys will perform much better than ARUs for estimating abundance of certain species, especially blackbirds that occur at very high densities. A call index or a "too many to count" option for transcribers could make transcription more realistic for recordings with very high numbers of a particular species. For certain species, point counts, whether in-person or ARU, may be inappropriate, and other methods such as territory mapping may be required to obtain a density index. Our correction factors take these effects into account and will reduce systematic biases. Nevertheless, depending on the goal of a study, it may be good practice to do a visual survey of wetlands when deploying or retrieving ARUs as a safeguard for detecting individuals of species that may be missed entirely on recordings.

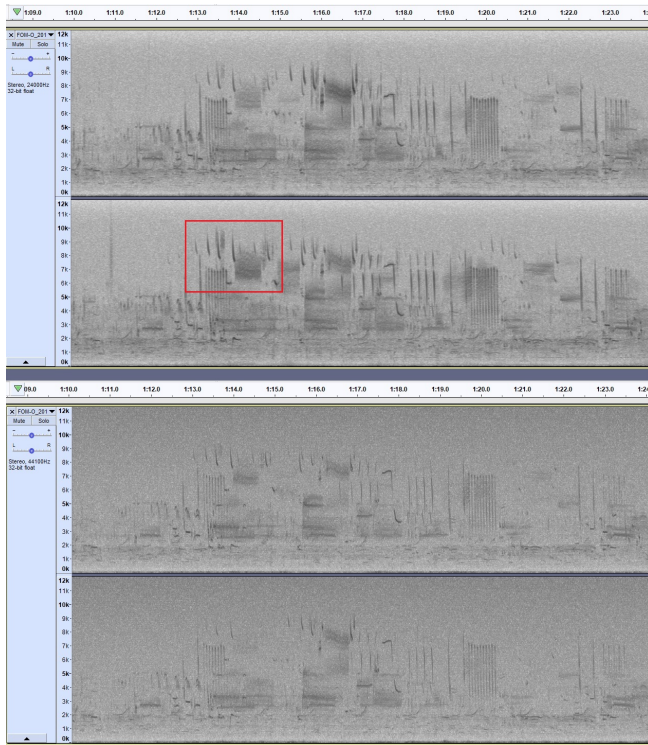
With the exception of Nelson's Sparrow, other species with low δ s for SM2s are species with low-pitched calls. Low-frequency environmental noise (wind, traffic), tends to be amplified on

recordings at lower frequencies (Fig. 6), possibly masking some of the songs of these species. This is similar to Bombaci and Pejchar (2019), who also found that species with low-pitched calls had lower δ s. However, Van Wilgenburg et al. (2017) found that the extremely low-pitched drumming of the Ruffed Grouse (*Bonasa umbellus*) was better detected by ARUs. The effect of wind was less pronounced on the SM4s than on the SM2s, and the species with low-pitched calls were generally detected just as well on SM4 recordings as by people in the field.

Species with positive biases and δ values larger than 1 were (Figs. 3 and 5): for SM2, Brown-headed Cowbird and Yellow Warbler; and for SM4, Sora, Savannah Sparrow, Yellow Warbler, Le Conte's Sparrow, Clay-colored Sparrow, and American Coot. Species with high-pitched calls (Brown-headed Cowbird, Savannah Sparrow, Le Conte's Sparrow) tend to be detected better by ARUs; perhaps these species are more difficult to hear in-person but more easily seen on a spectrogram. This effect is especially pronounced on SM4 recordings, which have more sensitive microphones, resulting in large differences in δ s for those species (Figs. 3 and 7).

Across species, the mean SM2 δ was slightly less than 1 (0.98) and estimates of δ were less than 1 for 14 of 19 species, supporting our prediction that τ_A would be slightly smaller than τ_H . However, the mean SM4 δ was slightly greater than 1 (1.03) and estimates of δ were greater than 1 for 10 of 16 species, indicating that τ_A s for SM4s are actually slightly larger than τ_H s. Measures of δ for SM4 recordings were higher than those for SM2 recordings for 15 of 16 species, supporting our prediction that τ_A would be larger for SM4s than for SM2s. As well, there are two species between our study and Van Wilgenburg et al. (2017) that provide opportunity for comparison. For Common Yellowthroat, our calculation of SM2 δ (0.98) was similar to that of Van Wilgenburg et al. (2017; $\delta = 1.00$). However, for Clay-colored Sparrow, our calculation was much larger ($\delta = 1.07$ compared to $\delta = 0.77$). Yip et al. (2017) found evidence for habitat-related variation in δ , and we speculate that the difference in Clay-colored Sparrow habitat in the boreal forest studied by Van Wilgenburg et al. (2017) versus the open prairie habitat in our study is responsible for this discrepancy.

Fig. 7. Spectrograms of simultaneous recordings at a wetland in Saskatchewan, Canada. Above, SM4; below, SM2. Higher microphone sensitivity (darker songs) and lower internal noise (lighter “background”) is evident on the SM4 spectrogram. The songs of several Savannah Sparrows (*Passerculus sandwichensis*) can be seen; the red box shows one as an example. The observer estimated two Savannah Sparrows in the field, four on the SM4, and three on the SM2.



There are several types of ARUs, as well as several high-quality nonautonomous recording units that are affordable to hobbyists and are useful in situations when a person is available to operate the unit during recording. Rempel et al. (2013) tested several older models of recording units and showed that different makes and models had different microphone sensitivities and signal-to-noise ratios. They also showed that both signal-to-noise ratios and microphone sensitivities varied with audio frequency and that units with lower signal-to-noise ratios detected fewer birds compared to human surveyors and more sensitive recording units. Similarly, Darras et al. (2018) found by meta-analysis of 23 published papers that signal-to-noise ratio of microphones had a positive effect on species richness. Thus, each model of recording unit with its particular specifications will need to be tested to determine the most accurate model- and species-specific correction factors.

Our estimates of availability (p) were close to 1 for most songbirds, especially for Red-winged and Yellow-headed Blackbirds, which we encountered at high densities in Saskatchewan. By contrast, availability estimates were lower and associated uncertainty was large for more elusive marsh birds, e.g., grebes, Virginia Rail, Sora, and American Coot. This is likely due to less-frequent singing in

these species (Conway 2011). These species account for examples where there was somewhat weak correlation between estimates of p_H and p_A . Despite low correlation, p_A was higher than p_H for these species, even in the context of a paired point count, with the exception of American Coot and Virginia Rail (SM2). The low availability of elusive marsh birds is precisely why call-broadcast is typically used for in-person surveys, because it entices the birds to sing and therefore increases availability. However, call-broadcast affects species movement and biases population estimates (Conway and Gibbs 2011, Zuberogitia et al. 2011). The use of ARUs can overcome low availability without using call-broadcast by transcribing multiple recordings from the same station, i.e., repeat visits, to increase overall probability of detection. Algorithms can be used to estimate the number of recordings that would need to be processed in order to have a good chance of detecting a particular species (Sliwinski et al. 2016). Using the algorithms of Sliwinski et al. (2016), to have > 90% chance of detecting a species when $p = 0.28$ (our lowest p_A estimate, for Virginia Rail) would require at least seven 5-min surveys. The transcription of this many-repeat visits would increase the cost of manual transcription, but would ultimately provide more accurate population estimates, and could be more cost-effective than whatever number of repeat in-person visits might be required (Conway and Gibbs 2011).

For this suite of elusive marsh birds, it is possible that availability will be higher in the absence of a human observer. Studies on forest bird populations have shown that a human observer has no effect on position of individual birds, probability of occurrence, or singing rate (Campbell and Francis 2012, Hutto and Hutto 2020), and we expect similar results for many of our species. However, changes in marsh bird behavior due to the presence of an observer is an area needing more controlled research. A methodology similar to that of Campbell and Francis (2012), who used microphone arrays to test movement and vocalization of upland birds before and after the arrival of an observer, could be applied to marsh birds. This could be an important area of research for a suite of so-called “secretive” species for which call-broadcast surveys are advocated to increase availability in real time during the course of a survey. If p_A is indeed different when a human is not present, this could be accounted for in any offsets in order to integrate in-person and ARU point counts. Because many elusive marsh birds have highest detectability at night (Tozer et al. 2016, 2017), and because most in-person counts are conducted at dawn or dusk, differences in availability at different times of day will also need to be included into statistical offsets for both ARU and in-person counts.

Most estimates of p_H are similar to those of p_A ; however, p_A tended to be slightly higher and have smaller error than p_H , even for elusive marsh birds with generally low p (Fig. 2). This was also found by Van Wilgenburg et al. (2017; see Fig 2, which shows more species above the 1:1 correspondence line than below) and Bombaci and Pejchar (2019; see Table 1). In time removal models, shorter times to first detection lead to higher measures of p . Therefore, a higher p is equivalent to a shorter average time to first detection. This means that observers detected most birds slightly earlier on recordings than they did in the field, likely because of the ability to pause or rewind a recording to verify identifications. This pattern even occurred to some extent with grebes, suggesting that even without visual cues observers may be

able to detect them earlier on a recording. The pattern also applies to both species that failed the equal availability test: Song Sparrow and Brown-headed Cowbird. We speculate that observers in the field needed to listen to the Song Sparrow several times to eliminate similar-sounding species (e.g., Vesper Sparrow, *Poocetes gramineus*). Similarly, the call of the Brown-headed Cowbird is weak, very high, and short; therefore, it may be easy to miss in the field during the first few minutes when observers are focused on the louder and more prominent marsh birds. On the other hand, this call shows up very well on a spectrogram, and with the ability to rewind observers may be less overwhelmed in a lab setting. For these two species and other species that do not have equal availability between survey types, e.g., Ruffed Grouse (Van Wilgenburg et al. 2017), *p* could be incorporated into statistical offsets as was done by Sólymos et al. (2013) for point counts of varying lengths and at different times of day and year.

ARUs are flexible and effective tools that offer a wide range of benefits. A recorder deployed for the full spring-summer season in temperate North America is able to collect data on species that vocalize in early spring, e.g., amphibians or owls, early summer, e.g., grebes, dawn (most songbirds), dusk, e.g., nightjars, and night, e.g., rails and bitterns (Shonfield and Bayne 2017). Autonomous recorders will address many challenges of marsh bird monitoring such as infrequent singing and reluctance of elusive species to vocalize when a human is present (Conway and Gibbs 2011). Bird song identification skills are not necessarily needed by field personnel who deploy and retrieve the recorders, thereby greatly expanding the pool of potential qualified field personnel, including volunteers for citizen science monitoring programs (Dickinson et al. 2010). With the emerging consensus that data from recorders are comparable to human observations, and the fact that methodological biases can be understood and accounted for (Sólymos et al. 2013, Van Wilgenburg et al. 2017, Yip et al. 2017, this study), we recommend that ARUs see increasing use to facilitate delivery of wetland bird surveys.

Responses to this article can be read online at:
<http://www.ace-eco.org/issues/responses.php/1661>

Acknowledgments:

We acknowledge Laura Tranquilla for support of field work and the field technicians who collected data for the project: Enid Cumming, JJ Mackenzie, LeeAnn Latremouille, Laura Achenbach, Amanda Guercio, Chris Ketola, Tim Arthur, Jeremy Bensette, Paulette Herbert, and Laura Cardenas-Ortiz. We thank Steve Van Wilgenburg for guidance in data analysis. The project was supported by Birds Canada's Long Point Waterfowl and Wetlands Research Program, Environment and Climate Change Canada, SC Johnson, and The Bluff's Hunting Club. And finally, we thank Keith Hobson, two anonymous reviewers, and members of the Scientific Advisory Committee of Birds Canada's Long Point Waterfowl and Wetlands Research Program for comments that improved the paper.

LITERATURE CITED

Akaike, H. 1973. Information theory and the maximum likelihood principle. In B.N. Petrov and F. Csäki, editors. 2nd

International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Alquezar, R. D., and R. B. Machado. 2015. Comparisons between autonomous acoustic recordings and avian point counts in open woodland savanna. *Wilson Journal of Ornithology* 127:712-723. <https://doi.org/10.1676/14-104.1>

Beier, P., K. P. Burnham, and D. R. Anderson. 2001. Model selection and inference: a practical information-theoretic approach. *Journal of Wildlife Management*. [online] URL: https://www.mendeley.com/catalogue/38785021-22d0-34b4-b81b-3968a3fd0163/?utm_source=desktop&utm_medium=1.19.4&utm_campaign=open_catalog&userDocumentId=%7B8090f3ff-60cf-4b5e-afc6-6a1b506a55b8%7D

Bobay, L. R., P. J. Taillie, and C. E. Moorman. 2018. Use of autonomous recording units increased detection of a secretive marsh bird. *Journal of Field Ornithology* 89:384-392. <https://doi.org/10.1111/jfo.12274>

Bombaci, S. P., and L. Pejchar. 2019. Using paired acoustic sampling to enhance population monitoring of New Zealand's forest birds. *New Zealand Journal of Ecology* 43:3356. <https://doi.org/10.20417/nzj ecol.43.9>

Borker, A. L., P. Halbert, M. W. Mckown, B. R. Tershy, and D. A. Croll. 2015. A comparison of automated and traditional monitoring techniques for Marbled Murrelets using passive acoustic sensors. *Wildlife Society Bulletin* 39:813-818. <https://doi.org/10.1002/wsb.608>

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction of distance sampling: estimating abundance of biological populations*. Oxford University Press, Oxford, UK.

Campbell, M., and C. M. Francis. 2012. Using microphone arrays to examine effects of observers on birds during point count surveys. *Journal of Field Ornithology* 83:391-402. <https://doi.org/10.1111/j.1557-9263.2012.00389.x>

Conway, C. J. 2011. Standardized North American marsh bird monitoring protocol. *Waterbirds* 34:319-346. <https://doi.org/10.1675/063.034.0307>

Conway, C. J., and J. P. Gibbs. 2011. Summary of intrinsic and extrinsic factors affecting detection probability of marsh birds. *Wetlands* 31:403-411. <https://doi.org/10.1007/s13157-011-0155-x>

Dalh, T. E. 2006. *Status and trends of wetlands in the conterminous United States 1986 to 1997*. U.S. Fish and Wildlife Service, Onalaska, Wisconsin, USA.

Darras, K., P. Batáry, B. Furnas, A. Celis-Murillo, S. L. Van Wilgenburg, Y. A. Mulyani, and T. Tschardtke. 2018. Comparing the sampling performance of sound recorders versus point counts in bird surveys: a meta-analysis. *Journal of Applied Ecology* 55:2575-2586. <https://doi.org/10.1111/1365-2664.13229>

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter. 2010. Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics* 41:149-172. <https://doi.org/10.1146/annurev-ecolsys-102209-144636>

Drake, K. L., M. Frey, D. Hogan, and R. Hedley. 2016. Using digital recordings and sonogram analysis to obtain counts of

- Yellow Rails. *Wildlife Society Bulletin* 40:346-354. <https://doi.org/10.1002/wsb.658>
- Farnsworth, G. L., K. H. Pollock, J. D. Nichols, T. R. Simons, J. E. Hines, and J. R. Sauer. 2002. A removal model for estimating detection probabilities from point-count surveys. *Auk* 119:414-425. <https://doi.org/10.1093/auk/119.2.414>
- Haselmayer, J., and J. S. Quinn. 2000. A comparison of point counts and sound recording as bird survey methods in Amazonian southeast Peru. *Condor* 102:887-893. [https://doi.org/10.1650/0010-5422\(2000\)102\[0887:ACOPCA\]2.0.CO;2](https://doi.org/10.1650/0010-5422(2000)102[0887:ACOPCA]2.0.CO;2)
- 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. <https://doi.org/10.2307/3784223>
- Hutto, R. L., and R. R. Hutto. 2020. Does the presence of an observer affect a bird's occurrence rate or singing rate during a point count? *Journal of Field Ornithology* 91:214-223. <https://doi.org/10.1111/jfo.12329>
- Klingbeil, B. T., and M. R. Willig. 2015. Bird biodiversity assessments in temperate forest: the value of point count versus acoustic monitoring protocols. *PeerJ* 3:e973. <https://doi.org/10.7717/peerj.973>
- Matsuoka, S. M., E. M. Bayne, P. Sólymos, P. C. Fontaine, S. G. Cumming, F. K. A. Schmiegelow, and S. J. Song. 2012. Using binomial distance-sampling models to estimate the effective detection radius of point-count surveys across boreal Canada. *Auk* 129:268-282. <https://doi.org/10.1525/auk.2012.11190>
- Rempel, R. S., C. M. Francis, J. N. Robinson, and M. Campbell. 2013. Comparison of audio recording system performance for detecting and monitoring songbirds. *Journal of Field Ornithology* 84:86-97. <https://doi.org/10.1111/jfo.12008>
- Rempel, R. S., J. M. Jackson, S. L. Van Wilgenburg, and J. A. Rodgers. 2019. A multiple detection state occupancy model using autonomous recordings facilitates correction of false positive and false negative observation errors. *Avian Conservation and Ecology* 14(2):1. <https://doi.org/10.5751/ACE-01374-140201>
- Saunders, S. P., K. A. L. Hall, N. Hill, and N. L. Michel. 2019. Multiscale effects of wetland availability and matrix composition on wetland breeding birds in Minnesota, USA. *Condor* 121:duz024. <https://doi.org/10.1093/condor/duz024>
- Sedláček, O., J. Vokurková, M. Ferenc, E. N. Djomo, T. Albrecht, and D. Hořák. 2015. A comparison of point counts with a new acoustic sampling method: a case study of a bird community from the montane forests of Mount Cameroon. *Ostrich* 86:213-220. <https://doi.org/10.2989/00306525.2015.1049669>
- Shonfield, J., and E. M. Bayne. 2017. Autonomous recording units in avian ecological research: current use and future applications. *Avian Conservation and Ecology* 12(1):14. <https://doi.org/10.5751/ACE-00974-120114>
- Sidie-Slettedahl, A. M., K. C. Jensen, R. R. Johnson, T. W. Arnold, J. E. Austin, and J. D. Stafford. 2015. Evaluation of autonomous recording units for detecting 3 species of secretive marsh birds. *Wildlife Society Bulletin* 39:626-634. <https://doi.org/10.1002/wsb.569>
- Sliwinski, M., L. Powell, N. Koper, M. Giovanni, and W. Schacht. 2016. Research design considerations to ensure detection of all species in an avian community. *Methods in Ecology and Evolution* 7:456-462. <https://doi.org/10.1111/2041-210X.12506>
- Snell, E. A. 1987. *Wetland distribution and conversion in southern Ontario*. Inland Waters and Lands Directorate, Environment Canada, Burlington, Ontario, Canada.
- Sólymos, P., S. M. Matsuoka, E. M. Bayne, S. R. Lele, P. Fontaine, S. G. Cumming, D. Stralberg, F. K. A. Schmiegelow, and S. J. Song. 2013. Calibrating indices of avian density from non-standardized survey data: making the most of a messy situation. *Methods in Ecology and Evolution* 4:1047-1058. <https://doi.org/10.1111/2041-210X.12106>
- Sólymos, P., M. Moreno, and S. R. Lele. 2018. *detect: Analyzing wildlife data with detection error*. R package version 0.4-2.
- Steer, J. 2010. Bioacoustic monitoring of New Zealand birds. *Notornis* 57:75-80.
- Steidl, R. J., C. J. Conway, and A. R. Litt. 2013. Power to detect trends in abundance of secretive marsh birds: effects of species traits and sampling effort. *Journal of Wildlife Management* 77:445-453. <https://doi.org/10.1002/jwmg.505>
- Tiner, R. W. 1984. *Wetlands of the United States: current status and recent trends*. National Wetlands Inventory, Fish and Wildlife Service, U.S. Department of the Interior, Washington, D.C., USA.
- Tozer, D. C. 2016. Marsh bird occupancy dynamics, trends, and conservation in the southern Great Lakes basin: 1996 to 2013. *Journal of Great Lakes Research* 42:136-145. <https://doi.org/10.1016/j.jglr.2015.10.015>
- Tozer, D. C., K. F. Abraham, and E. Nol. 2006. Improving the accuracy of counts of wetland breeding birds at the point scale. *Wetlands* 26:518-527. [https://doi.org/10.1672/0277-5212\(2006\)26\[518:ITAOCO\]2.0.CO;2](https://doi.org/10.1672/0277-5212(2006)26[518:ITAOCO]2.0.CO;2)
- Tozer, D. C., K. L. Drake, and C. M. Falconer. 2016. Modeling detection probability to improve marsh bird surveys in southern Canada and the Great Lakes states. *Avian Conservation and Ecology* 11(2):3. <https://doi.org/10.5751/ACE-00875-110203>
- Tozer, D. C., C. Myles Falconer, A. M. Bracey, E. E. Gnass Giese, G. J. Niemi, R. W. Howe, T. M. Gehring, and C. J. Norment. 2017. Influence of call broadcast timing within point counts and survey duration on detection probability of marsh breeding birds. *Avian Conservation and Ecology* 12(2):8. <https://doi.org/10.5751/ace-01063-120208>
- Turgeon, P. J., S. L. Van Wilgenburg, and K. L. Drake. 2017. Microphone variability and degradation: implications for monitoring programs employing autonomous recording units. *Avian Conservation and Ecology* 12(1):9. <https://doi.org/10.5751/ACE-00958-120109>
- Van Wilgenburg, S. L., P. Sólymos, K. J. Kardynal, and M. D. Frey. 2017. Paired sampling standardizes point count data from humans and acoustic recorders. *Avian Conservation and Ecology* 12(1):13. <https://doi.org/10.5751/ACE-00975-120113>

Vold, S. T., C. M. Handel, and L. B. McNew. 2017. Comparison of acoustic recorders and field observers for monitoring tundra bird communities. *Wildlife Society Bulletin* 41:566-576. <https://doi.org/10.1002/wsb.785>

Wildlife Acoustics. 2011. Song Meter SM2 User Manual. Wildlife Acoustics, Maynard, MA.

Wildlife Acoustics. 2019. *Song Meter SM4 User Guide*. Wildlife Acoustics, Maynard, Massachusetts, USA.

Yip, D. A., L. Leston, E. M. Bayne, P. Sólymos, and A. Grover. 2017. Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conservation and Ecology* 12(1):11. <https://doi.org/10.5751/ACE-00997-120111>

Znidarsic, E., M. Towsey, W. K. Roy, S. E. Darling, A. Truskinger, P. Roe, and D. M. Watson. 2020. Using visualization and machine learning methods to monitor low detectability species—the Least Bittern as a case study. *Ecological Informatics* 55:101014. <https://doi.org/10.1016/j.ecoinf.2019.101014>

Zuberogoitia, I., J. Zabala, and J. E. Martínez. 2011. Bias in little owl population estimates using playback techniques during surveys. *Animal Biodiversity and Conservation* 34(2).

Zwart, M. C., A. Baker, P. J. K. McGowan, and M. J. Whittingham. 2014. The use of automated bioacoustic recorders to replace human wildlife surveys: an example using nightjars. *PLoS ONE* 9:e102770. <https://doi.org/10.1371/journal.pone.0102770>



Appendix 1.

Table A1.1. Sample sizes for each focal species. Sample sizes represent number of surveys where each species was detected.

Species	Latin Name	Alpha Code	Paired SM2		Paired SM4	
			In-person	SM2	In-person	SM4
Pied-billed Grebe	<i>Podilymbus podiceps</i>	PBGR	93	81	52	51
Red-necked Grebe	<i>Podiceps grisegena</i>	RNGR	35	27	27	27
Virginia Rail	<i>Rallus limicola</i>	VIRA	41	22	8	6
Sora	<i>Porzana carolina</i>	SORA	128	129	76	85
Common Gallinule	<i>Gallinula galeata</i>	COGA	6	4	0	0
American Coot	<i>Fulica americana</i>	AMCO	16	14	15	17
Wilson's Snipe	<i>Gallinago delicata</i>	WISN	94	97	76	84
American Bittern	<i>Botaurus lentiginosus</i>	AMBI	62	40	19	14
Least Bittern	<i>Ixobrychus exilis</i>	LEBI	17	7	0	0
Sedge Wren	<i>Cistothorus platensis</i>	SEWR	43	37	44	42
Marsh Wren	<i>Cistothorus palustris</i>	MAWR	49	52	39	43
Clay-colored Sparrow	<i>Spizella pallida</i>	CCSP	38	35	41	45
LeConte's Sparrow	<i>Ammospiza leconteii</i>	LCSP	44	41	47	56
Nelson's Sparrow	<i>Ammospiza nelsoni</i>	NESP	43	36	8	11
Savannah Sparrow	<i>Passerculus sandwichensis</i>	SAVS	96	107	73	101
Song Sparrow	<i>Melospiza melodia</i>	SOSP	140	146	112	120
Swamp Sparrow	<i>Melospiza georgiana</i>	SWSP	21	22	24	23
Yellow-headed Blackbird	<i>Xanthocephalus xanthocephalus</i>	YHBL	24	22	24	27
Red-winged Blackbird	<i>Agelaius phoeniceus</i>	RWBL	49	50	51	53
Brown-headed Cowbird	<i>Molothrus ater</i>	BHCO	25	41	26	36
Common Yellowthroat	<i>Geothlypis trichas</i>	COYE	73	78	75	80
Yellow Warbler	<i>Setophaga petechia</i>	YEWA	26	27	28	29

Table A1.2. Maximum counts of each focal species obtained on a point count for each survey type.

Species	Paired SM2		Paired SM4	
	In-person	SM2	In-person	SM4
Pied-billed Grebe	3	3	2	2
Red-necked Grebe	8	4	8	3
Virginia Rail	2	1	--	--
Sora	5	4	3	4
American Coot	2	1	2	2
Wilson's Snipe	4	6	4	3
American Bittern	4	3	2	2
Sedge Wren	3	3	3	3
Marsh Wren	8	4	4	4
Clay-colored Sparrow	4	4	4	4
LeConte's Sparrow	3	3	3	3
Nelson's Sparrow	5	3	--	--
Savannah Sparrow	3	4	3	4
Song Sparrow	5	3	5	4
Swamp Sparrow	3	4	3	3
Yellow-headed Blackbird	30	10	30	10
Red-winged Blackbird	50	10	50	10
Brown-headed Cowbird	5	3	5	5
Common Yellowthroat	4	3	4	3
Yellow Warbler	3	4	3	4

Table A1.3. Effective detection radii for in-person counts and 95% confidence intervals.

Species	LCL	EDR_H (m)	UCL
Pied-billed Grebe	152	181	216
Red-necked Grebe	131	172	226
Virginia Rail	71	84	101
Sora	111	124	138
American Coot	82	112	153
Wilson's Snipe	123	143	165
American Bittern	121	144	171
Sedge Wren	67	79	93
Marsh Wren	62	70	79
Clay-colored Sparrow	66	77	90
LeConte's Sparrow	44	51	60
Nelson's Sparrow	71	81	92
Savannah Sparrow	56	62	68
Song Sparrow	66	71	77
Swamp Sparrow	67	83	102
Yellow-headed Blackbird	84	110	143
Red-winged Blackbird	83	97	113
Brown-headed Cowbird	68	85	105
Common Yellowthroat	58	65	72
Yellow Warbler	73	87	104

Table A1.4. Availability estimates for in-person (p_H) and ARU (p_A) paired surveys. Availability was calculated using maximum likelihood estimates of singing rates. Bold denotes species where survey type was found to have a significant effect on availability.

Species	Paired SM2 surveys						Paired SM4 surveys					
	LCL	p_H	UCL	LCL	p_A	UCL	LCL	p_H	UCL	LCL	p_A	UCL
Pied-billed Grebe	0.09	0.34	0.84	0.31	0.58	0.87	0.01	0.21	1	0.26	0.58	0.93
Red-necked Grebe	0.06	0.37	0.96	0.5	0.77	0.95	0.13	0.48	0.95	0.23	0.58	0.94
Virginia Rail	0.17	0.53	0.96	0	0.28	1	0.81	0.9	0.96	0.87	0.93	0.97
Sora	0.54	0.7	0.84	0.5	0.67	0.83	0.47	0.68	0.88	0.58	0.75	0.89
American Coot	0.79	0.95	1	0.17	0.67	1	0.7	0.93	1	0.36	0.76	0.99
Wilson's Snipe	0.87	0.93	0.97	0.91	0.96	0.98	0.81	0.9	0.96	0.87	0.93	0.97
American Bittern	0.77	0.89	0.96	0.76	0.9	0.98	0.19	0.64	0.99	0.51	0.86	1
Sedge Wren	0.89	0.96	0.99	0.82	0.93	0.99	0.9	0.96	0.99	0.96	0.99	1
Marsh Wren	0.94	0.97	0.99	0.98	0.99	1	0.97	0.99	1	0.99	1	1
Clay-colored Sparrow	0.93	0.97	0.99	0.98	0.99	1	0.93	0.97	0.99	0.98	0.99	1
LeConte's Sparrow	0.87	0.95	0.99	0.88	0.96	0.99	0.87	0.95	0.99	0.96	0.99	1
Nelson's Sparrow	0.57	0.77	0.92	0.76	0.9	0.98						
Savannah Sparrow	0.93	0.96	0.98	0.95	0.97	0.99	0.91	0.96	0.98	0.97	0.99	1
Song Sparrow	0.74	0.83	0.91	0.91	0.95	0.97	0.73	0.83	0.91	0.94	0.97	0.99
Swamp Sparrow	0.9	0.97	1	0.98	1	1	0.92	0.98	1	0.99	1	1
Yellow-headed Blackbird	1	1	1	1	1	1	1	1	1	1	1	1
Red-winged Blackbird	1	1	1	1	1	1	1	1	1	1	1	1
Brown-headed Cowbird	0.01	0.29	1	0.84	0.94	0.99	0	0	0.01	0.75	0.9	0.98
Common Yellowthroat	0.86	0.93	0.98	0.94	0.98	0.99	0.9	0.96	0.98	0.95	0.98	0.99
Yellow Warbler	0.96	0.99	1	0.99	1	1	0.96	0.99	1	0.97	0.99	1

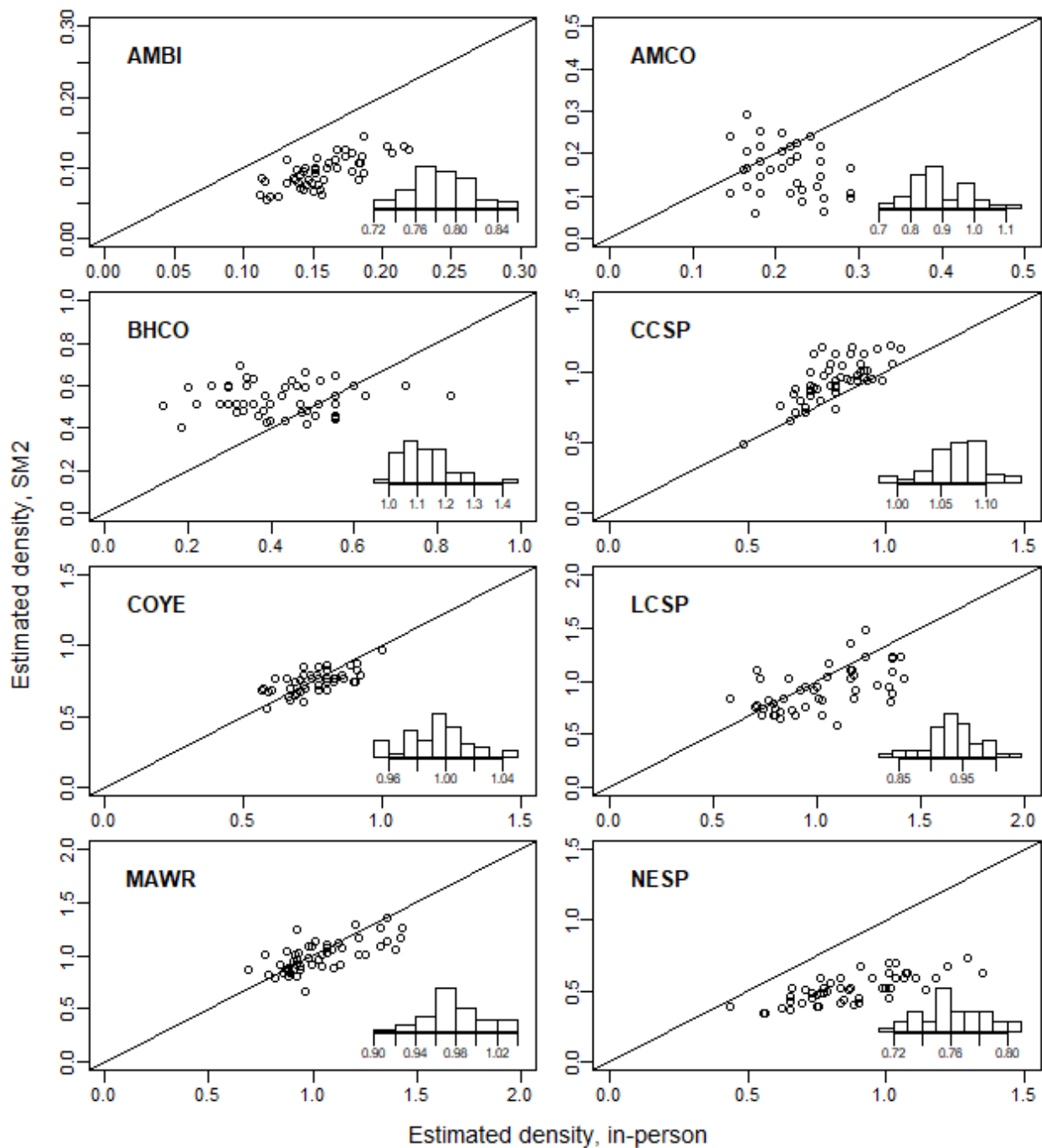
Table A1.5. Estimated correction factors (δ) for SM2 surveys, with 95% confidence intervals. Maximum Likelihood Estimates are given, as well as results of calibration and validation values from repeated random subsampling. For details of calculations, see Methods.

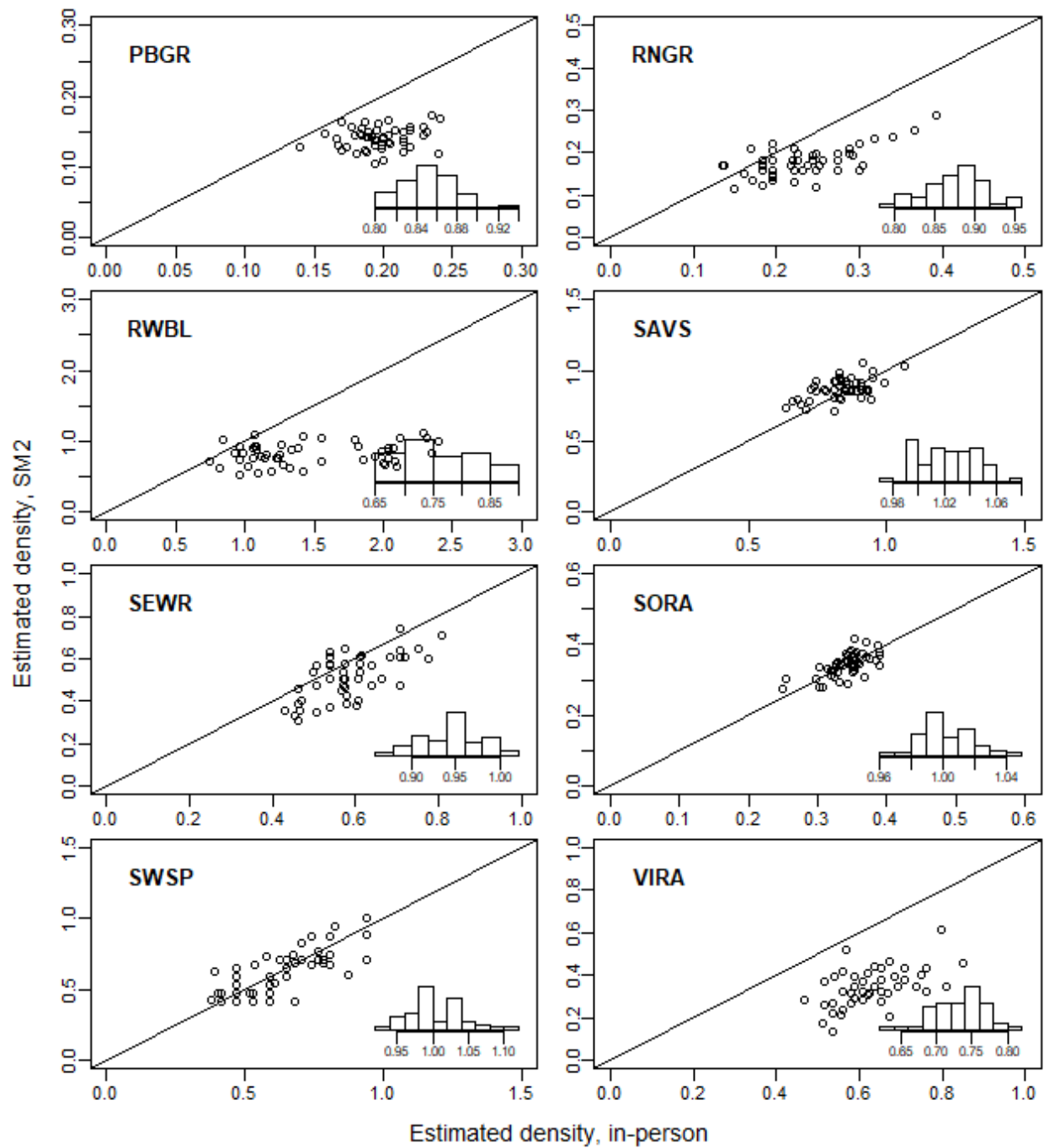
Species	MLE Estimates			Repeated Random Sampling					
				Calibration			Validation		
	LCL	δ	UCL	LCL	δ	UCL	LCL	δ	UCL
Pied-billed Grebe	0.75	0.86	0.99	0.81	0.85	0.9	0.74	0.85	0.97
Red-necked Grebe	0.75	0.9	1.08	0.8	0.88	0.96	0.73	0.89	1.12
Virginia Rail	0.5	0.68	0.93	0.66	0.73	0.79	0.56	0.73	0.88
Sora	0.91	1	1.1	0.98	1	1.04	0.92	1	1.08
American Coot	0.65	0.92	1.3	0.74	0.89	1.09	0.58	0.89	1.27
Wilson's Snipe	0.86	0.98	1.1	0.96	0.99	1.02	0.92	0.99	1.07
American Bittern	0.67	0.81	0.99	0.73	0.78	0.84	0.67	0.77	0.88
Sedge Wren	0.81	0.98	1.18	0.89	0.95	1.01	0.82	0.94	1.06
Marsh Wren	0.85	0.98	1.12	0.93	0.98	1.04	0.87	0.99	1.14
Clay-colored Sparrow	0.9	1.07	1.25	1	1.07	1.12	0.98	1.07	1.22
LeConte's Sparrow	0.79	0.94	1.13	0.85	0.93	1.01	0.78	0.96	1.2
Nelson's Sparrow	0.64	0.77	0.91	0.72	0.76	0.81	0.67	0.77	0.89
Savannah Sparrow	0.91	1.02	1.15	0.99	1.02	1.06	0.94	1.02	1.1
Swamp Sparrow	0.79	0.99	1.26	0.94	1	1.09	0.83	1	1.16
Yellow-headed Blackbird	0.68	0.8	0.94	0.64	0.78	1.01	0.51	0.85	1.24
Red-winged Blackbird	0.68	0.76	0.85	0.68	0.76	0.88	0.57	0.77	1
Brown-headed Cowbird	0.91	1.13	1.4	1.02	1.12	1.27	0.9	1.17	1.69
Common Yellowthroat	0.86	0.98	1.13	0.96	1	1.04	0.92	1	1.1
Yellow Warbler	0.92	1.1	1.31	1.05	1.09	1.14	1	1.1	1.2

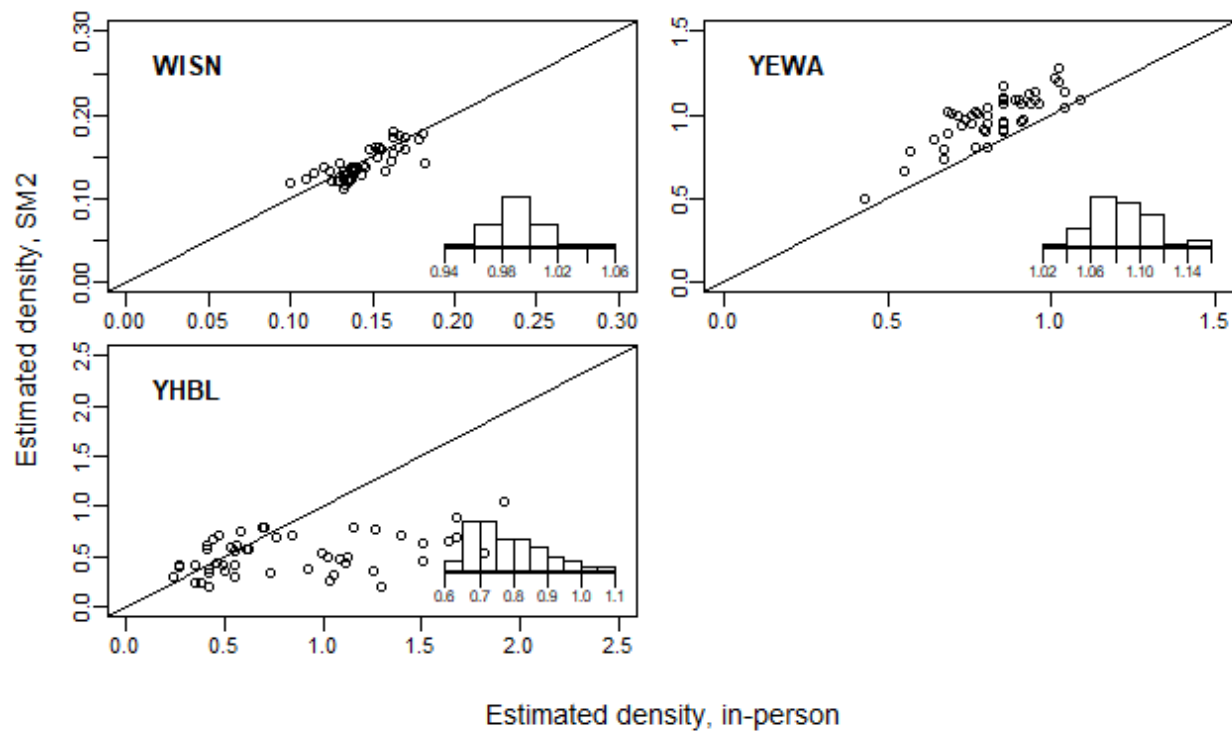
Table A1.6. Estimated correction factors (δ) for SM4 surveys, with 95% confidence intervals. Maximum Likelihood Estimates are given, as well as results of calibration and validation values from repeated random subsampling. For details of calculations, see Methods.

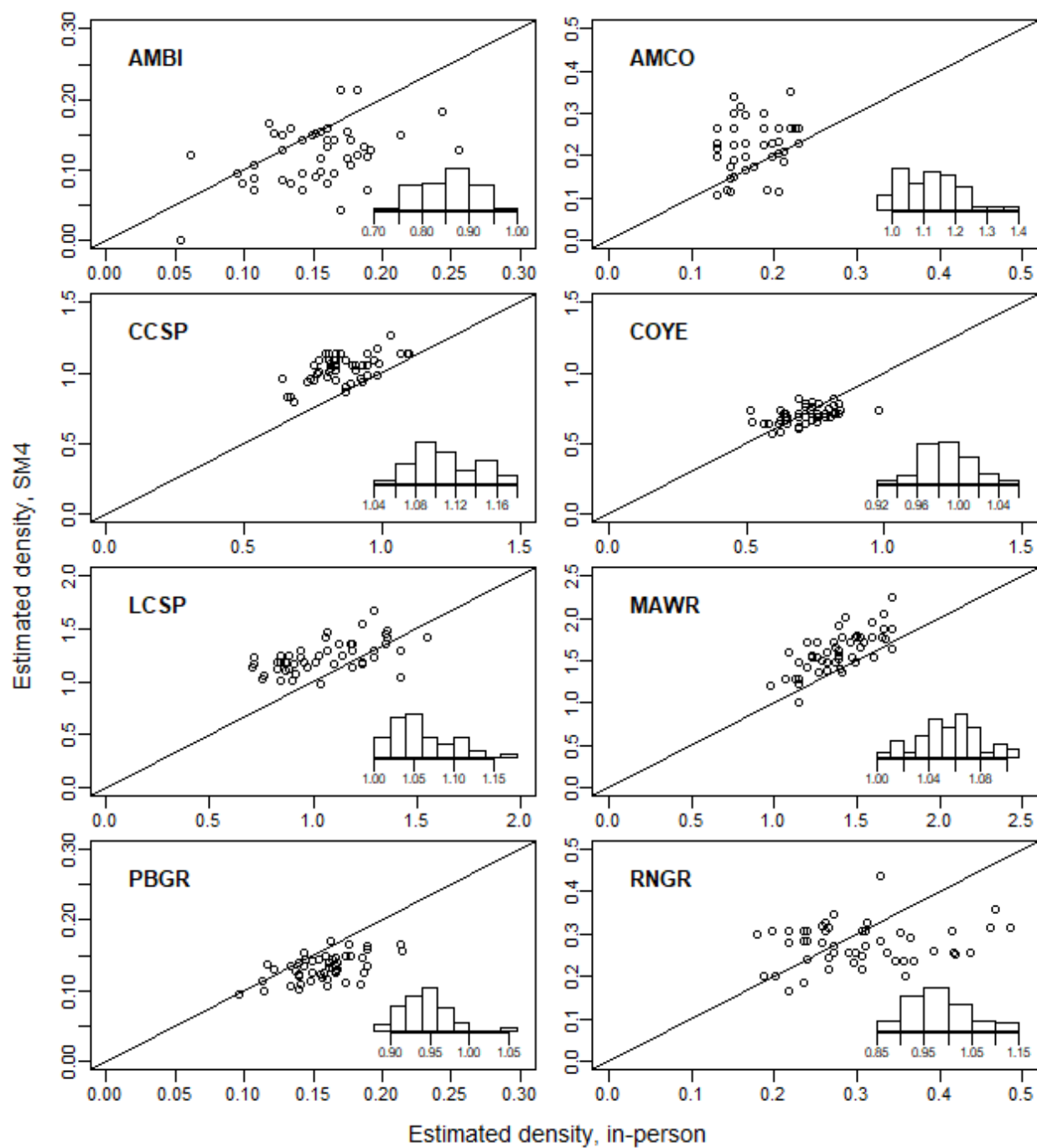
Species	MLE Estimates			Repeated Random Sampling					
	LCL	δ	UCL	Calibration			Validation		
				LCL	δ	UCL	LCL	δ	UCL
Pied-billed Grebe	0.82	0.97	1.15	0.9	0.94	1	0.8	0.92	1.04
Red-necked Grebe	0.81	0.98	1.2	0.89	0.98	1.12	0.77	0.97	1.23
Sora	1	1.13	1.27	1.05	1.08	1.11	0.98	1.06	1.14
American Coot	0.8	1.1	1.52	1	1.12	1.32	0.81	1.14	1.41
Wilson's Snipe	0.9	1.03	1.17	0.99	1.03	1.06	0.95	1.04	1.12
American Bittern	0.63	0.87	1.19	0.76	0.86	0.93	0.53	0.88	1.17
Sedge Wren	0.87	1.04	1.25	0.97	1.03	1.11	0.91	1.02	1.15
Marsh Wren	0.92	1.06	1.22	1.02	1.06	1.1	0.98	1.08	1.19
Clay-colored Sparrow	0.95	1.11	1.3	1.06	1.11	1.16	1	1.1	1.2
LeConte's Sparrow	0.92	1.08	1.29	1	1.06	1.12	0.96	1.09	1.28
Savannah Sparrow	1.05	1.18	1.32	1.14	1.19	1.25	1.07	1.19	1.32
Swamp Sparrow	0.75	0.95	1.2	0.88	0.95	1.02	0.77	0.94	1.14
Yellow-headed Blackbird	0.75	0.88	1.02	0.77	0.93	1.17	0.59	0.92	1.36
Red-winged Blackbird	0.71	0.78	0.87	0.69	0.81	0.96	0.56	0.8	1.09
Common Yellowthroat	0.88	1	1.15	0.95	0.99	1.02	0.92	1	1.11
Yellow Warbler	0.94	1.12	1.33	1.05	1.11	1.18	0.98	1.11	1.22

Fig. A1.1. In-person and uncorrected ARU estimates of species density (birds/ha). Calculations are based on GLM parameters of the 30% validation data subset within the repeated random subsampling framework. Each point represents one of 50 random subsamples. The line indicates 1:1 correspondence. Inset are histograms of GLM survey-effect estimates of δ using the 70% calibration data across all 50 subsamples. Where points tend to fall underneath the 1:1 line, δ is lower, and the larger the spread of the points the more uncertainty there is in the δ estimate. Figures are shown for each species and for both SM2 and SM4 densities. Graphs are labelled with 4-letter species codes (see Table A1).









Estimated density, SM4

