Combining point counts and autonomous recording units improves survey efficacy across elevational gradients on two continents

Anna Drake[1], Devin de Zwaan[2], Tomás Altamirano[3], Scott Wilson[4], Kristina Hick[5], Camila Bravo[6], JOSE IBARRA[7], and Kathy Martin[8]

[1]The University of British Columbia
[2]University of British Columbia Department of Forest and Conservation Sciences
[3]The University of British Columbia Department of Forest & Conservation Sciences
[4]Government of Canada
[5]Environment Canada Pacific and Yukon Region
[6]Universidad de Chile
[7]Pontificia Universidad Catolica de Chile
[8]University of British Columbia

February 11, 2021

Abstract

1. Accurate biodiversity and population monitoring is a requirement for effective conservation decision-making. Survey method bias is therefore a concern, particularly when research programs face logistical and cost limitations. 2. We employed point counts (PCs) and autonomous recording units (ARUs) to survey avian biodiversity across elevational gradients in comparable temperate mountain habitats at opposite ends of the Americas (9 mountains in British Columbia (BC), Canada and 10 in southern Chile). We compared detected species richness against multi-year species inventories and examined differences in detection probability by family. By incorporating time costs, we assessed the performance and efficiency of single vs. combined methods. 3. ARUs were predicted to capture ~92% of species present in BC but only ~58% in Chile, despite Chilean mountain communities being less diverse. Community, rather than landscape composition, appears to be the driver of this dramatic difference. Chilean communities contain less-vocal species, which ARUs missed. Further, 6/14 families in BC were better detected by ARUs while 11/11 families in Chile were better detected by PCs. Where survey conditions differentially impacted methods, PC detection varied over the morning and with canopy cover in BC and ARU detection probability mostly varied seasonally in Chile. Within a single year of monitoring, neither method alone was predicted to capture the full avian community, with the exception of ARUs in the alpine and subalpine of BC. PCs contributed little to detected diversity in BC, but including this method resulted in negligible increases in total time costs. Combining PCs with ARUs in Chile significantly increased species detections, again, for little cost. 4. Combined methods were among the most efficient and accurate approaches to capturing diversity. We recommend conducting observer point counts, where possible, when ARUs are deployed and retrieved, in order to capture additional diversity and flag methodology biases with minimal additional effort.

1

[2.] ECOS (Ecology-Complexity-Society) Laboratory, Center for Local Development (CEDEL) & Center for Intercultural and Indigenous Research (CIIR), Villarrica Campus, Pontificia Universidad Católica de Chile.

[3.] Environment and Climate Change Canada, Wildlife Research Division, National Wildlife Research Centre, Ottawa, ON, Canada.

[4.] Environment and Climate Change Canada, Pacific Wildlife Research Centre, Vancouver, BC, Canada.

[5.] Institue of Ecology and Biodiversity, Department of Ecological Sciences, Faculty of Science, Universidad de Chile, Santiago, Chile.

[6.] Millennium Nucleus Center for the Socioeconomic Impact of Environmental Policies (CESIEP) & Center of Applied Ecology and Sustainability (CAPES), Pontificia Universidad Católica de Chile

*Corresponding author: drake.ae@gmail.com

ABSTRACT

1. Accurate biodiversity and population monitoring is a requirement for effective conservation decision-making. Survey method bias is therefore a concern, particularly when programs face logistical and cost limitations.
2. We employed point count (PCs) and autonomous recording unit (ARUs) to temperate mountain habitats (9 mountains in British Columbia (BC), Canada and 10 in southern Chile). We compared detected species richness against multi-year species inventories and examined differences in detection probability by family. By incorporating time costs, we assessed the performance and efficiency of single vs. combined methods.
3. ARUs were predicted to capture ~92% of species present in BC but only ~58% in Chile, despite Chilean mountain communities being less diverse. Community, rather than landscape composition, appears to be the driver of this dramatic difference. Chilean communities contain less-vocal species, which ARUs missed6/14 families in BC were better detected by ARUs while 11/11 families in Chile were better detected by PCs. Where survey conditions differentially impacted methods, PC detection varied over the morning and with canopy cover in BC and ARU detection probability mostly varied seasonally in Chile. Within a single year of monitoring, neither method alone was predicted to capture the full avian community, with the exception of ARUs in the alpine and subalpine of BC. PCs contributed little to detected diversity in BC, but including this method increases total time costs. Combining PCs with ARUs in Chile significantly increased species detections, again, for little cost.
4. Combined methods were among the most efficient and accurate approaches to capturing diversity. We recommend conducting observer point counts, where possible, when ARUs are deployed and retrieved, in order to capture additional diversity and flag methodology biases with.

KEY-WORDS

autonomous recording units, point count surveys, avian diversity, biodiversity monitoring, elevational gradient, species richness, high mountain biodiversity, alpine

Species surveys are used to determine the presence, relative abundance, and diversity of taxa over space and time (Roberts, 2011; Sauer et al., 2017; Schramm et al., 2020). cornerstone of ecological studies used to identify biodiversity hotspots, infer the impact of natural or anthropogenic disturbance on communities, assess the effectiveness of management practices, and identify important habitats for species of conservation concern (e.g. Dorji et al., 2019; Friedlander et al., 2019; Ibarra & Martin, 2015; Rosenberg et al., 2017). For effective conservation decision-making to occur, biases associated with any given survey technique should be quantified and, where possible, corrected for. When abundance and diversity data are compared across broad regions and divergent communities, any interaction between detection bias due to survey method and the landscapes and/or communities being surveyed is a concern. The use of survey method can reveal such problems and may increase project coverage and efficiency.

For birds in terrestrial habitats, point counts have been the standard survey method for more than 80 years (Ralph et al., 1995). Point counts employ 1-2 trained observers to identify and count birds by sight and sound from a single location for a set period of time. Within the past 20 years, the use of autonomous recording units (ARUs) as an alternative to point count surveys has become increasingly popular (Darras et al., 2019). ARUs are installed at survey sites and record ambient sound that is then in the lab, with species identified by their vocalizations either manually or using identification software. Both methods have benefits and limitations as techniques for surveying avian diversity. Key among the benefits of point counts is the ability to visually species (Acevedo & Villanueva-Rivera, 2006; Hutto & Stutzman, 2009; Vold et al., 2017) and use distance to obtain better density estimates than can be assessed by audio alone (Shonfield & Bayne, 2017). Because point countobservers can assess call direction and they outperform ARUs when calls occur outside the ARU microphone(s) "line-of-sight" (Castro et al., 2019). ARUs, on the other hand, overcome logistical constraints experienced by point counts that can impact species detections. ARUs can collect data simultaneously from multiple sites, allowing projects to survey during peak diel activity for both diurnal and nocturnal species (Goyette et al., 2011) and eliminating potential time bias present in point counts (Darras et al., 2019). ARUs can be left in high latitude and high elevation habitats year-round and programed to start recording in spring, before observers can access these regions (e.g. Shonfield & Bayne, 2017). They can, therefore, better-sample peak seasonal activity for resident species and detect shifts in bird phenology (Klingbeil & Willig, 2015). Finally, as inanimate objects, ARUs are less likely to alter bird behaviour (Shonfield & Bayne, 2017, Darras et al., 2019, *but see* Hutto & Hutto 2020.

Effort is a consideration for research programs. Point counts and ARUs differ in their time costs. A single point count is completed in a single site visit. Establishing an ARU site and collecting data entails a minimum of two site visits; however, ARU recordings can subsequently be intensively sampled without increased field costs or increased site disturbance. ARUs can have notable drawbacks in terms of processing time in the lab: without automated data processing, the time costs of uploading and interpreting audio files, replaying sections of audio, and then transcribing observations is greater than for detections and transcriptions of equivalent length point counts (e.g. this study; Celis-Murillo et al., 2009; Alquezar & Machado, 2015). Even with automated processing, the need to proof output can eliminate any time advantages over manual scanning (Joshi et al., 2017 *but see* Knight et al., 2020).

Despite the fact that mountains support important bird diversity, most high elevation systems in the Americas are poorly monitored (Boyle & Martin, 2015). Mountain habitats present challenging conditions in which to conduct avian surveys. Access is often limited by difficult terrain, late snowmelt, and poor infrastructure. Surveys may be disrupted by extreme weather. By necessity, mountain surveys are typically conducted in a linear fashion upslope or downslope, creating time bias in point counts across elevation. Given their field advantages, ARUs offer a compelling alternative to point counts at high elevation sites. Here we examine the performance of ARUs and point count surveys in detecting and quantifying avian diversity across a gradient of temperate mountain habitats in both North and South America. In both countries sampling encompassed three structurally similar habitats: densely forested upper montane, open subalpine, and highly exposed alpine. Using species detections at shared sites, we directly compare diversity index values and species accumulation curves produced by these two methods. We investigate the underlying causes of differences in diversity values obtained by each method by modeling method effects on the detection probabilities of bird families within the two regions. In order to make recommendations for future monitoring protocols, we examine the efficiency (time cost versus species detection return) of point counts and ARU sampling on their own, and for combined-method protocols.

MATERIALS & METHODS

Study locations

In Canada, in 2019, we surveyed nine mountains in the D'ze Kant Bulkley Valley) and Kitimat-Stikine regions of British Columbia (BC; 1000-1801 m elevation). In Chile, in 2018, we surveyed 10 mountains in La Araucanía and Los Ríos regions (1000-1700 m elevation). These sites fall within the traditional unceded lands of the Wet'suwet'en, Gitxsan, and Tsimshian First Nations in BC and the Mapuche peoples in Chile.

The farthest Boyle & Martin, 2015).

BC survey sites fall within biogeoclimatic zonesMountain Hemlock, Engelmann Spruce-Subalpine Fir, Boreal Altai Fescue Alpine,nd Coastal Mountain-heatherBritish Columbia Ministry of Forests, Lands, Natural Resource Operations, and Rural Development, 2018). Montane habitat is primarily old growth . The subalpine consists of woody shrubs, grasses and perennial herbs with some tree cover; the alpine is characterized by the presence of fescue grasses, mosses, and lichens.

In Chile, montane habitats are dominated by old growth mixed broadleaf-conifer forests, with about 10 % mid-successional forest. Subalpine habitat is a mix of highland herbaceous meadows, shrubs, and sparse patches of trees and/or krummholz. Perennial herbaceous plants, shrubs, few or no trees, and bare rock/scree characterize alpine habitat. Vegetation structure varies within- and among-mountains based on natural disturbance (i.e. volcanic eruptions) and/or land use history (Caviedes & Ibarra, 2017).

Point Counts

Surveys started at sunrise and ran until 1000to encompass peak bird activity. Each mountain was surveyed using transects with five point counts in each of the three habitat types for a total of 15 point counts per mountain. Point counts were separated by 200 m to avoid repeat counts of birds. During each 6-minute point count, bird species and abundance was assessed by sight and sound with distance from the count centre recorded at four intervals (0-25 m, 26-50 m, 51-100 m, >100m). Counts were repeated three times: between May 30 – July 16 in BC, and between November 7 – December 21 in Chile to assess detection probability and address seasonal variation in detection.

Acoustic recordings and analysis

Song Meter SM4 Autonomous Recording Units (ARUs; Wildlife Acoustics Inc. ©, Maynard, MA) with two omni-directional microphones were deployed at two point count sites per habitat (six per mountain) in both BC and Chile. ARUs were deployed at least two point count sites apart (>400 m). In BC, 36 units were deployed for 10 – 25 days on six mountains, and for 33 – 37 days on the remaining 3 mountains, between June 3 and July 15. In Chile, 6 units were deployed on five mountains for 3 – 10 days each, between November 13 – December 28. ARUs recorded at a sampling rate of 24000 Hz in stereo wav format using default acoustic gain settings for the microphones. Units were mounted on a tree several meters from the point count site, or on a PVC pipe at ˜1.5 m height in the alpine. ARUs were programmed to record 30-mins on, 30-min off starting three hours before sunset, all night, and ending 5 hours after sunrise (16 x 30-minutes point counts).

Sound recordings were analyzed using Audacity® software (V2.3.0, Audacity Team, 2020). Two skilled observers reviewed all the recordings in BC and bserver did so in Chile. All observers had experience conducting point counts in the same regions. In BC, both observers analyzed five of the same recordings to confirm detection consistency. Spectrograms were scanned manually in stereo format as the observer listened to the recording. Species that were more difficult to identify were compared with recordings available on bioacoustics libraries such as the Xeno-Canto Foundation (2019) and/or sent to other skilled ornithologists.

Two (BC) or three (Chile) different days were selected for sampling per site over the season. In BC, . If a selected day had unfavorable conditions (wind or rain) that interfered with the audio, another day was selected randomly. From the 30-minute audio recorded hourly between sunrise and four hours after sunrise on these days, we randomly chose one, 6-minute interval to analyze (5 x 6-minute point counts/day). We manually analyzed 1,150 x 6-minute ARU counts.

Abiotic Variables

During each point count, we collected abiotic information within a 50 m buffer of the point count site. We collected temperature and wind speed using Kestrel 3500 weather meters (Nielsen-Kellerman Company, PA, USA). We also recorded wind as a categorical variable (0-3). Additionally, we recorded percent canopy, understory (vegetation ˜30cm in height), shrub, tundra, snow, and rock cover and the percentage of the ground covered with dead trees. Deciduous trees were more common in the canopy in Chile than in BC.

4

Canopy cover values in Chile therefore increased with leaf-out during the season while values in BC were static.

For each ARU count we also scored wind on the Beaufort scale (0 – 3) based on its interference with the audio recording.

Statistical Analyses

Minimum species richness by habitat

We derived a minimum species richness for avian communities in each habitat using the most complete species list compiled between 2017-2019 at our field sites. This list included species identified at point counts, while walking transects between point count sites (K. Martin et al.,), as well as species identified in this study's ARU recordings. These values therefore represent the minimum known species richness for each community.

Species diversity indices by survey method

All analyses were completed R (R Core Team, 2019). For diversity indices, we restricted our datasets to sites that were surveyed by both ARUs and PC methods (BC: n=52 sites, Chile: n=30 sites). We then produced species accumulation curves for each method, using species incidence frequencies and the program iNEXT (Hsieh et al. 2016). For ARUs, within-day hourly measures (dawn - 5 hrs after) were sampled independently (BC: n=44-47 site-surveys/habitat/hour; Chile: n=30 site-surveys/habitat/hour) and were also pooled (BC: n=220-236 site-surveys/habitat; Chile: n=150 site-surveys/habitat) for direct comparison with point count survey data (BC: n=48-54 site-surveys/ habitat; Chile: n=30 site-surveys/habitat). Sample sizes are larger for BC because we had access to more ARUs (see above). In both BC and Chile, diversity indices were calculated for each accumulation curve at 97% sample completeness throughinterpolation/extrapolation. This allowed for a fair comparison of the performance of each method (and each time period within ARU counts) regardless of sample size/effort. We report two diversity metrics (Hill numbers): richness (q=0) and the effective number of species calculated by the exponential of the Shannon-Wiener Index (q=1), plus their 84% CI (MacGregor-Fors & Payton, 2013) (Fig. 1). Richness is presented as the count of species captured by either method. The exponential Shannon-Wiener value weights species by their frequency of occurrence and therefore the importance of species detected only once or twice by either method.

We used the ChaoRichness function in iNext to predict the asymptote of the species richness accumulation curves of each method (Chao, 1984). This value is the predicted final species richness detected by each method if effort was increased. We compare these values to our minimum species diversity in each habitat.

Detection probability by method

For species that were detected by one method only, we assessed the probability that this was due to a detection difference between methods versus chance using the Fisher's exact test on the frequency of detection by method across all site-surveys (Fisher, 1992).

Because we were also interested in generalizable patterns of detection, we pooled species into family groups and assessed each family's detection probability, by method, using the R package, 'unmarked (Fiske & Chandler, 2011; Table S2(A) and S2(B)). For point counts, detections at all sites were used for modelling detection probability, including sites that did not have ARUs installed (BC: n=129 sites; Chile: n=150 sites). The number of repeated surveys at each site ranged from 3 (point count only sites) to 23 (sites with both ARU and point count data) (BC: n=1065 site-surveys; Chile: 900 site-surveys). We restricted the families modelled to those that occupied at least 15% of sites in any of our three habitat types. Detection modeling was restricted to those habitats where 90% of occupied sites occurred.

Because ARUs were repeatedly sampled within-day with spacing of ~ 1 hour (58 ±13 min), we expected temporal autocorrelation between surveys within-site and incorporated this into our models using a first-order Markov covariate (Wright et al., 2016).

Our base detection probability model was:

detection ~ wind score + hours after sunrise + hours after sunrise^2 + date + date^2 + canopy cover + canopy cover^2 + temporal autocorrelation term

And site occupancy probability was modeled as:

occupancy ~ site elevation + residuals of canopy cover by elevation.

Canopy cover residuals were used in the occupancy model to account for co-linearity between elevation and canopy (i.e. trees become more sparse at higher elevations). In Chile, canopy cover values at the time of sampling were used for modeling detection to account for leafing-out, while maximum canopy cover at each site (reflective of habitat type) was used for modeling occupancy.

To our base detection model, we added an effect of method (ARU vs. PC) on detection plus interactions between method and 3 survey parameters where effects on detection were predicted to differ between ARU and point counts. These were: canopy cover, hours after sunrise, and date. We tested the performance of the basic model, the basic + method model, and the seven possible models that included combinations of 'method x survey condition'. In total, nine detection models were tested for each bird family.

We selected the best model for each family based on QAIC, incorporating ĉ for the most complex model (detection ~ basic model + method + all three method interactions) (Burnham & Anderson, 2002, MacKenzie et al., 2017; Mazerolle, 2017). Goodness-of-fit tests were run for these best models and, where ĉ > 1, we inflate the CIs accordingly. We do not present output for any family where ĉ > 4 (suggesting lack of fit; Mazerolle, 2017) or where ĉ < 0.3 (indicating insufficient data). We report the 84% and 95% CIs: no overlap at the 84% CI is consistent with a significant difference (P<0.05) between methods (Payton et al., 2003) while the 95% CI represents the 95% CI of the actual detection probability. Further detail on detection probability modelling is available in the Supplement.

e assessed the efficiency of single-method and mixed-method sampling protocols as the percent of the total community detected as a function of hours of effort. For ARUs, site visitation and sample processing cost was assessed at 40 min/site and 9 min/sample. For point counts, these values were 20 min/site and 7 min/sample. When protocols were mixed, we assumed that the visitation cost was shared for ARUs and PCs, i.e. that point counts were conducted when ARUs were deployed and/or retrieved. In protocols that involved 3 point counts per site, the additional point count incurred an additional visitation cost (20 min/site). We randomly sampled ARU and point count surveys with replacement (10,000X) at each survey site to produce a bootstrapped mean species richness detected (±SE) across all sites for different sampling intensities of: ARUs alone (1-15 counts/site), point counts alone (1-3 counts/site), and point count plus ARU surveys (1 point count plus 1-15 ARU counts/site, 2 point counts plus 1-15 ARU counts/site, etc.). We identify the "best" protocols as those that detected the greatest percentage of the total community for the least effort.

RESULTS

Species diversity indices

In BC, at 97% predicted community coverage, point counts and pooled ARUs obtained equivalent species richness (q=0) in both the alpine and the subalpine (Fig. 1). Pooled ARUs obtained higher richness scores than point counts in the upper montane. When species were weighted by their frequency of occurrence in either dataset (q=1), the methods performed equivalently in the subalpine, but pooled ARUs outperformed point counts in the alpine and upper montane (Fig. 1). Thus 3 of 6 comparisons in BC showed equivalent performance for the two methods and 3 indicated ARUs were superior, particularly in the upper montane.

In BC, ARU detections were more likely to intersect with our lower estimate of the entire community than point counts in the alpine and subalpine (Fig. 2). In the upper montane, predicted final species richness for both methods did not approach our lower estimate of the entire community. On average, pooled ARUs were predicted to capture 92% (range: 83-100%) of the entire community across all habitats. Point counts were predicted to capture 73% (62-79%).

6

In Chile, at 97% predicted community coverage, point counts obtained greater species richness (q=0) values than pooled ARUs in the subalpine and alpine (Fig. 1). In upper montane forest, the richness obtained by both methods was equivalent. When first order diversity (q=1) was assessed, point counts continued to be better than pooled ARUs in detecting species diversity, outperforming ARUs in the upper montane as well. Thus 5 out of 6 comparisons in Chile indicated that point counts outperformed ARUs; the sixth showed a bias toward point counts, but was non-significant (Fig. 1).

For both methods in Chile, the predicted asymptotes of the species accumulation curves did not approach our lower estimate of the entire community (Fig. 2). On average, pooled ARUs were predicted to capture 58% (range: 50-68%) of the entire community across all habitats. Point counts were predicted to capture 70% (57-78%).

In both regions, over the morning, dawn ARU counts detected lower or equivalent richness to counts later in the morning (q=0, Fig. 1). The only exception was in the BC alpine, where dawn counts detected more species than counts two hours after dawn (q=0, Fig. 1). Although dawn recordings were less rich, in Chile they detected two owl and one nightjar species that were not detected later in the morning (*see* Table S1(B) and below).

Species identified by only one method

In BC, ARUs detected all of the species by point count plus an additional 16, or 28% of the total diversity detected by both methods (Table S1(A)). Of these species, only Townsend's solitaire (*Myadestes townsendi*) was detected frequently enough by ARUs to indicate that the detection difference between methods was not due to chance (Fisher's exact test; P=0.02).

In Chile, 13 species, or 26% of the diversity captured by both methods, were detected by point count observers but missed by ARUs (Table S1(B)). Most of these were raptors (6/13) and ground-tyrants (5/13; Tyrannidae). Of these 13 species, 5 were detected frequently enough by point counts to indicate that the detection difference between methods was not a product of chance (Fisher's exact test; P<0.05). These species were: Bar-winged cinclodes (*Cinclodes fuscus),* Dark-faced ground-tyrant (*Muscisaxicola maclovianus*), Spot-billed ground-tyrant (*M. maculirostris)* , and Red-backed hawk (*Geranoetus polyosoma)* . species, or 8% of the diversity captured by both methods, were detected by ARUs but missed by point counts (Table S1(B)). All of these species were nocturnal and all, with the exception of the Austral pygmy owl (*Glaucidium nana*), were detected only in dawn ARU recordings. None were detected frequently enough to exclude the possibility that the detection difference between methods was due to chance.

Family-level detection probabilities by method

In BC, models supported an effect of methodology on detection for only 6 of the 14 families examined. These were: finches (Fringillidae), sparrows (Passerellidae), warblers (Parulidae), kinglets (Regulidae), thrushes (Turdidae) and corvids (Corvidae; Figs S1-S3). All six families were better detected by ARUs. Kinglets, warblers and thrushes showed an interaction between methodology and hours after sunrise: detection probability declined over the morning for point counts but remained consistently high for ARUs (Fig. S2). Warblers, thrushes and corvids showed an interaction between methodology and canopy cover: detection probability was more variable over the range of canopy cover for point counts than for ARUs (Fig. S3).

In Chile, detection models for all 11 families examined supported a methodology effect, with a higher detection probability in point counts than in ARUs (Figs. S4-S6). Of these, six families showed an interaction between methodology and date: ARU detection probability for swallows (Hirundinidae), hummingbirds (Trochilidae), woodpeckers (Picidae), wrens (Troglodytidae), and parakeets (Psittacidae was either lower early in the monitoring period or exhibited a mid-season dip. The detectability of ovenbirds (Furnariidae), showed a mid-season dip in point counts but not ARUs (Fig. S4). Ovenbirds additionally had lower detection probability for ARUs under conditions of high canopy cover. Swallows were better detected by point counts in the mid-morning; wrens were more poorly detected by ARUs in the early morning (Fig. S5).

Protocol efficiency and performance comparisons

7

In BC, species accumulation as a function of hours of effort was indistinguishable between ARU-only protocols and one- and two-point count rounds plus ARU sampling (Fig. 3A). This was because point counts did not contribute the total survey cost but, as shown above, they also did not contribute novel species to the accumulation curve. Three-point count rounds and a mixed method that included point counts at this intensity were the least-efficient sampling protocol in BC due to the increased cost associated with a third site visit. Surprisingly, in BC, a single ARU count/site detected more species than two point counts/site in the subalpine and more than three point counts/site in the alpine and upper montane, for less effort (13 vs. 14 hrs and 15 vs. 24 hrs, respectively; Fig. 3A).

In Chile, ARUs alone were less efficient than point counts alone and less efficient than mixed methods due to the fewer species detections acquired by ARUs. This was particularly notable in the Chilean alpine, where a single point count/site detected more species than 10 ARU counts/site and two point counts/site detected more species than were detected at our maximum ARU effort of 15 counts/site, for less effort (4.5 vs. 22 hrs and 9 vs. 29 hrs, respectively; Fig. 3B). While species accumulation curves of mixed methods showed a large degree of overlap, a minimum of two point counts/site supplemented with ARUs appeared to be the best methodology for the subalpine and upper montane in Chile, and three point counts/site in the alpine boosted species detections enough to warrant the additional visitation cost (Fig. 3B).

DISCUSSION

Avian surveys using ARUs can overcome major limitations experienced by point count methods. In our study system, these include limitations associated with remote, difficult terrain and late snowmelt as well as the disruption of surveys due to inclement weather. Such advantages potentially make ARUs a powerful substitute for point counts (e.g. Darras et al., 2019). Our results here, however, indicate that ARUs should be augmented by point counts: dual methods allowed us to identify detection differences between methods where they were not anticipated. In this specific case, performance differences are likely attributable to differences in community composition between regions (as we discuss below). More generally however, our results show how dual methods enable monitoring programs to flag detection issues associated with survey method and thus enhance comparisons across habitat types and ecosystems.

High mountain habitats in BC and Chile are structurally similar, yet ARU performance was markedly better in BC than in Chile. This illustrates that avian community composition can matter as much as habitat composition in influencing method performance. As in Klingbeil and Willig (2015), we believe differences in detection probability that favour point counts in Chile are largely due to visual identification of species rather than audio detection. Raptor diversity is higher in Chile than BC and this largely silent group is best monitored by point counts. ARUs missed 6 raptor species that were picked up by point counts (Table S1). Similarly, Tyrannidae) rarely vocalize: the Xeno-canto Foundation notes that, of all neotropical genera, ground-tyrants and shrike-tyrants are the difficult to record. 5/9 tyrant species recorded in this study were missed by ARUs. Changes in vocalization frequency may also drive the seasonal variation in ARU detectability observed for 5/11 families in Chile. Song activity likely wanes when females are incubating or when pairs are feeding young (Moussus et al., 2009); yet, these individuals may remain visible during point counts when foraging. Interestingly, seasonal variation in detection probability was not supported for any family in BC.

ARUs provide the ability to re-play audio in order to capture all calls and confirm species identity. In contrast, point counts are more vulnerable to observer effects: individuals at point counts may miss species because they subconsciously screen out certain calls ("window species"; Kepler & Scott, 1981), are overwhelmed with the number of calling species (Celis-Murillo et al., 2009; Hutto & Stutzman, 2009), or because they mis-identify difficult calls (Bart, 1985; Celis-Murillo et al., 2009). This may explain why ARUs perform well in the species-rich upper montane (Fig. 1), and why a single ARU count/site in BC detected more species than a single point count/site, despite observation effort being equivalent (6 min/site; Fig. 2A and Fig. 3A). Two alternative explanations - that ARUs capture species' peak activity because they sample a broader period of the morning, or that ARUs fail to screen out songs originating outside of their focal habitat and therefore overstate species diversity - were not well supported by our data. First, richness by hour showed no evidence

of a peak in BC (Fig. 1A). Neither was there an ARU detection peak over the morning within-families (Fig. S2). Warblers, thrushes and kinglets were all, however, less likely to be detected by point counts later in the morning, pointing toward observer bias in point counts (Fig. S2). Secondly, as vocalizations tend to carry upslope, we would expect ARUs near habitat transition zones to mis-assign species to higher elevation habitats. Instead, ARUs in BC detected greater species diversity than point counts in upper montane habitat, not in the subalpine or alpine (Fig. 1A).

The ability to collect large amounts of data from ARUs is one of their advantages and, because the collection process itself is cheap, there is a temptation to obtain as much data as possible. However, the added time cost per sample associated with processing ARU data, when compared to point count surveys, needs to be carefully considered when planning monitoring protocols. Advances in automated processing may change this calculation (e.g. Knight et al., 2020), but additional time costs associated with training algorithms and proofing output still need to be considered (Joshi et al., 2017; Knight et al., 2017). Where ARUs perform poorly, as in the mountains of southern Chile, repeated sampling does not improve survey coverage (Fig. 2B). In other words, ARUs, like point counts, may miss large portions of communities regardless of effort. Programs should ascertain if this is the case before investing in increased ARU sampling. In this study, increased effort involved increased sampling within-day: it is possible that sampling more days, with lower effort within-day, would yield better returns. Detections of four nocturnal species in dawn ARU recordings highlight the benefit of synchronous sampling across survey sites.

Our work aligns with smaller studies that conclude dual methods are advantageous across a range of habitats (Celis-Murillo et al., 2009; 2012 (in specific cases); Tegeler et al., 2012; Alquezar & Machado, 2015; Vold et al., 2017), as well as two larger studies within temperate and boreal forest (Holmes et al., 2014; Van Wilgenburg et al., 2017). Our comparison across structurally similar habitats in different geographic regions highlights the importance of the avian community, in addition to habitat, in impacting method performance. We additionally show that the benefit-to-time-cost ratio of dual methods that employ 1-2 point counts/site is comparable or better than single-method approaches. Because our study system has relatively low species richness, our time costs for ARU transcription is relatively short. Where ARU processing is more time consuming, the benefits of employing dual methods should be more pronounced.

CONCLUSIONS

Because point counts survey sites differently than ARUs, their dual employment can identify situations where the performance of either method is strong or weak. When site visitation costs are shared, dual-method surveys are efficient and can markedly increase community coverage. We therefore recommend that, where possible, observer point counts be conducted when ARUs are deployed and when data are retrieved. Additionally, if some ARU recordings and point counts are conducted in tandem, point count data can be used to assess site-specific ARU detection radii (Van Wilgenburg et al., 2017; Yip et al., 2017). This would allow for better estimates of species densities from the audio data and help identify ARU species detection gaps (Vold et al., 2017). For occupancy studies, automated species detection software could then be used on longer sections of audio to efficiently search for species that have low ARU detection probabilities (Tegeler et al., 2012).

(74160073).

Data Accessibility. The datasets produced and analysed in this study are available at the Canadian Federal Open Government Portal [https://open.canada.ca/en/open-data; data DOI: *to be assigned when deposited*]

REFERENCES

Acevedo, M. A., & Villanueva-Rivera, L. J. (2006). Using Automated Digital Recording Systems as Effective Tools for the Monitoring of Birds and Amphibians. *Wildlife Society Bulletin* , *34(1)* , 211–214. https://doi.org/10.2193/0091-7648(2006)34[211:UADRSA]2.0.CO;2

Alquezar, R. D., & Machado, R. B. (2015). Comparisons Between Autonomous Acoustic Recordings and Avian Point Counts in Open Woodland Savanna. *The Wilson Journal of Ornithology* , *127(4)* , 712–723. https://doi.org/10.1676/14-104.1

Audacity Team. (2020). *Audacity(R): Free Audio Editor and Recorder* (Version 2.4.2) [Computer software]. https://audacityteam.org/

Bart, J. (1985). Causes of recording errors in singing bird surveys.*The Wilson Bulletin* , 161–172.

Boyle, W. A., & Martin, K. (2015). The conservation value of high elevation habitats to North American migrant birds. *Biological Conservation* , *192* , 461–476. https://doi.org/10.1016/j.biocon.2015.10.008

British Columbia Ministry of Forests, Lands, Natural Resource Operations, and Rural Development. (2018). *Biogeoclimatic Zones of British Columbia, 2018. Forest Analysis and Inventory Branch, Victoria, B.C. https://www.for.gov.bc.ca/hre/becweb/resources/* [Map]. https://www.for.gov.bc.ca/hre/becweb/resources/maps/index.html

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*(Second Edition). Springer-Verlag.

Castro, I., Rosa, A. D., Priyadarshani, N., Bradbury, L., & Marsland, S. (2019). Experimental test of birdcall detection by autonomous recorder units and by human observers using broadcast. *Ecology and Evolution* , *9(5)* , 2376–2397. https://doi.org/10.1002/ece3.4775

Caviedes, J., & Ibarra, J. T. (2017). Influence of Anthropogenic Disturbances on Stand Structural Complexity in Andean Temperate Forests: Implications for Managing Key Habitat for Biodiversity. *PLOS ONE* ,*12(1)* , e0169450. https://doi.org/10.1371/journal.pone.0169450

Celis-Murillo, A., Deppe, J. L., & Allen, M. F. (2009). Using soundscape recordings to estimate bird species abundance, richness, and composition. *Journal of Field Ornithology* , *80(1)* , 64–78. https://doi.org/10.1111/j.1557-9263.2009.00206.x

Celis-Murillo, A., Deppe, J. L., & Ward, M. P. (2012). Effectiveness and utility of acoustic recordings for surveying tropical birds.*Journal of Field Ornithology* , *83(2)* , 166–179. https://doi.org/10.1111/j.1557-9263.2012.00366.x

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics* , *11(4)* , 265–270. JSTOR.

Darras, K., Batary, P., Furnas, B. J., Grass, I., Mulyani, Y. A., & Tscharntke, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: A systematic map and user guide.*Ecological Applications* , *29(6)* , e01954. https://doi.org/10.1002/eap.1954

Dorji, S., Rajaratnam, R., & Vernes, K. (2019). Mammal richness and diversity in a Himalayan hotspot: The role of protected areas in conserving Bhutan's mammals. *Biodiversity and Conservation* ,*28(12)* , 3277–3297. https://doi.org/10.1007/s10531-019-01821-9

Fisher, R. A. (1992). Statistical Methods for Research Workers. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics: Methodology and Distribution* (pp. 66–70). Springer. https://doi.org/10.1007/978-1-4612-4380-9_6

Fiske, I., & Chandler, R. (2011). unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software* , *043* (i10). https://econpapers.repec.org/article/jssjstsof/v_3a043_3ai10.htm

Friedlander, A. M., Donovan, M. K., Koike, H., Murakawa, P., & Goodell, W. (2019). Characteristics of effective marine protected areas in Hawai'i. *Aquatic Conservation: Marine and Freshwater Ecosystems* , *29(S2)* , 103–117. https://doi.org/10.1002/aqc.3043

Goyette, J. L., Howe, R. W., Wolf, A. T., & Robinson, W. D. (2011). Detecting tropical nocturnal birds using automated audio recordings. *Journal of Field Ornithology* , *82(3)* , 279–287. https://doi.org/10.1111/j.1557-9263.2011.00331.x

Hobson, K. A., Rempel, R. S., Greenwood, H., Turnbull, B., & Van Wilgenburg, S. L. (2002). Acoustic Surveys of Birds Using Electronic Recordings: New Potential from an Omnidirectional Microphone System. *Wildlife Society Bulletin (1973-2006)* , *30(3)* , 709–720.

Holmes, S. B., McIlwrick, K. A., & Venier, L. A. (2014). Using automated sound recording and analysis to detect bird species-at-risk in southwestern Ontario woodlands. *Wildlife Society Bulletin* , *38(3)* , 591–598. https://doi.org/10.1002/wsb.421

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* , *7(12)* , 1451–1456. https://doi.org/10.1111/2041-210X.12613

Hutto, R. L., & Hutto, R. R. (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(2)* , 214–223. https://doi.org/10.1111/jofo.12329

Hutto, R. L., & Stutzman, R. J. (2009). Humans versus autonomous recording units: A comparison of point-count results. *Journal of Field Ornithology* , *80(4)* , 387–398. https://doi.org/10.1111/j.1557-9263.2009.00245.x

Ibarra, J. T., & Martin, K. (2015). Biotic homogenization: Loss of avian functional richness and habitat specialists in disturbed Andean temperate forests. *Biological Conservation* , *192* , 418–427. https://doi.org/10.1016/j.biocon.2015.11.008

Joshi, K. A., Mulder, R. A., & Rowe, K. M. C. (2017). Comparing manual and automated species recognition in the detection of four common south-east Australian forest birds from digital field recordings. *Emu - Austral Ornithology* , *117(3)* , 233–246. https://doi.org/10.1080/01584197.2017.1298970

Kepler, C. B., & Scott, J. M. (1981). Reducing Bird Count Variability by Training Observers. *Studies in Avian Biology* , *6* , 366–371.

Klingbeil, B. T., & Willig, M. R. (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

Knight, E. C., Sòlymos, P., Scott, C., & Bayne, E. M. (2020). Validation prediction: A flexible protocol to increase efficiency of automated acoustic processing for wildlife research. *Ecological Applications: A Publication of the Ecological Society of America* , e02140. https://doi.org/10.1002/eap.2140

Knight, E., Hannah, K., Foley, G., Scott, C., Brigham, R., & Bayne, E. (2017). Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology* , *12(2)* . https://doi.org/10.5751/ACE-01114-120214

Kułaga, K., & Budka, M. (2019). Bird species detection by an observer and an autonomous sound recorder in two different environments: Forest and farmland. *PLOS ONE* , *14(2)* , e0211970.

https://doi.org/10.1371/journal.pone.0211970

MacGregor-Fors, I., & Payton, M. E. (2013). Contrasting Diversity Values: Statistical Inferences Based on Overlapping Confidence Intervals. *PLOS ONE* , *8(2)* , e56794. https://doi.org/10.1371/journal.pone.0056794

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., & Hines, J. E. (2017). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence* . Elsevier.

Mazerolle, M. J. (2017). *Package "AICcmodavg": Model selection and multimodel inference based on (Q)AIC(c)* (2.2-2) [CRAN R Project].

Moussus, J.-P., Jiguet, F., Clavel, J., & Julliard, R. (2009). A method to estimate phenological variation using data from large-scale abundance monitoring programmes. *Bird Study* , *56(2)* , 198–212. https://doi.org/10.1080/00063650902792064

Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science* ,*3(1)* . https://doi.org/10.1093/jis/3.1.34

Ralph, C. J., John R. Sauer, & Sam Droege. (1995). *Monitoring Bird Populations by Point Counts* (Gen. Tech. Rep. PSW-GTR-149, p. 187). Pacific Southwest Research Station, Forest Service, US. Department of Agriculture.

Roberts, N. J. (2011). Investigation into survey techniques of large mammals: Surveyor competence and camera-trapping vs. transect-sampling.*Bioscience Horizons: The International Journal of Student Research* , *4(1)* , 40–49. https://doi.org/10.1093/biohorizons/hzr006

Rosenberg, K. V., Blancher, P. J., Stanton, J. C., & Panjabi, A. O. (2017). Use of North American Breeding Bird Survey data in avian conservation assessments. *The Condor* , *119(3)* , 594–606. https://doi.org/10.1650/CONDOR-17-57.1

Sauer, J. R., Pardieck, K. L., Ziolkowski, D. J., Smith, A. C., Hudson, M.-A. R., Rodriguez, V., Berlanga, H., Niven, D. K., & Link, W. A. (2017). The first 50 years of the North American Breeding Bird Survey.*The Condor* , *119(3)* , 576–593. https://doi.org/10.1650/CONDOR-17-83.1

Schramm, K. D., Harvey, E. S., Goetze, J. S., Travers, M. J., Warnock, B., & Saunders, B. J. (2020). A comparison of stereo-BRUV, diver operated and remote stereo-video transects for assessing reef fish assemblages. *Journal of Experimental Marine Biology and Ecology* ,*524* , 151273. https://doi.org/10.1016/j.jembe.2019.151273

Shonfield, J., & Bayne, E. (2017). Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology* , *12(1)* . https://doi.org/10.5751/ACE-00974-120114

Tegeler, A. K., Morrison, M. L., & Szewczak, J. M. (2012). Using extended-duration audio recordings to survey avian species.*Wildlife Society Bulletin* , *36(1)* , 21–29. https://doi.org/10.1002/wsb.112

Van Wilgenburg, S., Sólymos, P., Kardynal, K., & Frey, M. (2017). Paired sampling standardizes point count data from humans and acoustic recorders. *Avian Conservation and Ecology* , *12(1)* . https://doi.org/10.5751/ACE-00975-120113

Venier, L. A., Holmes, S. B., Holborn, G. W., Mcilwrick, K. A., & Brown, G. (2012). Evaluation of an automated recording device for monitoring forest birds. *Wildlife Society Bulletin* , *36(1)* , 30-39. https://doi.org/10.1002/wsb.88

Vold, S. T., Handel, C. M., & McNew, L. B. (2017). Comparison of acoustic recorders and field observers for monitoring tundra bird communities. *Wildlife Society Bulletin* , *41(3)* , 566–576. https://doi.org/10.1002/wsb.785

Wright, W. J., Irvine, K. M., & Rodhouse, T. J. (2016). A goodness-of-fit test for occupancy models with correlated within-season revisits. *Ecology and Evolution* , *6(15)* , 5404–5415. https://doi.org/10.1002/ece3.2292

Xeno-canto foundation. (2019). *Xeno-canto: Sharing bird songs from around the world* [www.xeno-canto.org].

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

FIGURES

FIGURE 1 Species diversity values (± 84% CI) obtained by point counters (blue points) and ARUs (pink and red points) across three mountain habitats in British Columbia and southern Chile. Values presented are species richness (Hill number (q) =0) and the effective number of species calculated by the exponential of the Shannon-Wiener Index (Hill number (q) =1). All values are interpolated/extrapolated to 97% sample completeness. Values for both hourly ARU counts (0-4 hrs after dawn) and full morning ARU data, pooled, are presented. The grey line between hourly points is a spline fit to aid in visualizing potential temporal trends. Significant differences between methods are indicated by a *

FIGURE 2 Species richness (q=0) accumulation curves for point counts (blue) and ARUs (red) across three montane habitats in British Columbia and in southern Chile (± 95% CI). Dashed lines indicate the predicted final species richness obtained by each method with increased effort. The solid orange line indicates the minimum community richness based on multi-year habitat sampling (*see* Methods).

FIGURE 3 Efficiency of single-method and dual-method protocols as the bootstrapped proportion of the community (mean ± SE) detected with increasing monitoring hours across mountain habitats in British Columbia (BC) and southern Chile. Species detections were summed across all survey sites (BC: n=16-18 sites/habitat, Chile: n=10 sites/habitat) for each level of effort. Point count returns (blue points) range in effort from 1-3 counts/site and are labelled. The ARU only protocol (red dashed curve) ranges in effort from 1-15 counts/site. Dual-method protocols (purple curves) range from 1-15 ARU counts/site and vary in point count effort as labelled.

Hosted file

Figure2_ARU.pdf available at https://authorea.com/users/395094/articles/508419-combining-point-counts-and-autonomous-recording-units-improves-survey-efficacy-across-elevational-gradients-on-two-continents

Hosted file

Figure4.pdf available at https://authorea.com/users/395094/articles/508419-combining-point-counts-and-autonomous-recording-units-improves-survey-efficacy-across-elevational-gradients-on-two-continents

(A) British Columbia

(B) Southern Chile