Received: 18 May 2021 | Revised: 3 June 2022 | Accepted: 9 June 2022

DOI: 10.1002/wsb.1370

TOOLS AND TECHNOLOGY

WILDLIFE SOCIETY BULLETIN

THE WILDLIFE SOCIETY

Acoustic detection of gunshots to improve measurement and mapping of hunting activity

Richard W. Hedley[1] | **Brian Joubert**[2] | **Harsimran K. Bains**[1] | **Erin M. Bayne**[1]

[1]University of Alberta, 11335 Saskatchewan Drive NW, Edmonton, AB T6G 2M9, Canada

[2]Alberta Environment and Parks, 9915 108 Street NW, Edmonton, AB T5K 2G6, Canada

Correspondence

Richard W. Hedley, University of Alberta, 11335 Saskatchewan Dr NW, Edmonton, AB, T6G 2M9, Canada.

Email: rhedley@ualberta.ca

Funding information

Mitacs; Natural Sciences and Engineering Research Council of Canada

Abstract

Hunting can influence the abundance and distribution of animals and act as a source of conflict among recreational user groups. Thus, land managers benefit from tools that can generate information about when and where hunting occurs. We used passive acoustic monitoring to examine spatio-temporal patterns of hunting-related gunshots at 91 locations in a protected area in Alberta, Canada. We compared 2 methods for detecting gunshots from recordings: a recognizer that used complex pattern recognition and an energy detector that detected loud sounds regardless of their acoustic features. The recognizer primarily detected faint sounds, and multiple observers showed low levels of agreement (37%) with respect to whether sounds were gunshots or not, suggesting it produced ambiguous data. The recognizer also missed many loud, clear gunshots for unknown reasons. The energy detector, in contrast, detected loud sounds upon which observers showed near-unanimous agreement (99%) on their identity. Gunshots missed by the energy detector could be because they were too quiet (i.e., too far away to be detected). Thus, despite detecting fewer gunshots overall, the energy detector produced higher quality data that were easier to interpret. We analyzed 249 gunshots detected with the energy detector, and found that hunting was concentrated near vehicle access points and peaked on Saturdays, and that hunters largely abided by local regulations prohibiting Sunday hunting. We compared energy detector results with remote cameras, which revealed similar spatiotemporal patterns of hunting effort. Passive acoustic monitoring has the potential to allow hunting activity to be mapped and monitored with unprecedented resolution.

KEYWORDS

acoustic monitoring, bioacoustics, gunshots, hunting, law enforcement monitoring, Song Meter SM4 Acoustic Recorder

Hunting poses unique challenges for land managers because of its effects on ecological systems and its role in human cultural systems. If left unchecked, excessive hunting can deplete wildlife populations, even in protected areas (Peres and Palacios 2007, Harrison 2011, Beirne et al. 2019). Too little hunting, on the other hand, can lead to undesirable densities of prey where natural predators are rare or absent (McShea 2012). Land managers are tasked with maintaining ecological balance while also considering a suite of other potential trade-offs when allowing a diversity of recreational activities, ensuring quality user experiences, and maintaining user safety (Tumes 2007). Navigating trade-offs requires detailed and accurate information on the rates at which users of protected areas engage in activities such as hunting to contextualize and understand its likely impacts on ecology and on other users.

Various methodologies have been employed to collect hunting-related data. Most commonly, hunter questionnaires have been used to study when and where hunting occurs (Bhandari et al. 2006, Peres and Palacios 2007, Stedman et al. 2008, Johnson et al. 2016). Less frequently, radio or global positioning system (GPS) collared animals have been tracked to examine spatial patterns of hunter-caused mortality (Fuller 1990, Stedman et al. 2004). Direct observations of hunters and evidence of hunting have also been used to quantify hunting pressure (Peres and Palacios 2007, Astaras et al. 2017). Each approach has its own strengths and weaknesses. For instance, hunter questionnaires can generate large data sets but rely on the cooperation of hunters and can therefore suffer from under-reporting, biased sampling, and potential for misleading data (Johnson et al. 2016). They also typically only reveal coarse spatial patterns of hunting activity (e.g., at the scale of a protected area; Peres and Palacios 2007). In Alberta, Canada, hunting effort is assessed annually via questionnaires that ask in which Wildlife Management Units an individual hunted, for how many days, and whether they were successful (Alberta Government 2021). Wildlife Management Units often span thousands of square kilometers, and hunting seasons last weeks or months. Thus, questionnaire information only reveals coarse details about hunter activities. Prey-focused studies that rely on collared animals can reveal fine-scale spatial patterns of mortality but require the deployment of a large number of collars (Fuller 1990) and may not show where hunters were most active, only where they were most successful. Meanwhile, attempts to directly detect hunters can be biased because some hunters—especially poachers—attempt to avoid detection (Burton et al. 2012, Astaras et al. 2017). Studies using camera traps as a mode of detection have reported that cameras were destroyed or stolen, which is likely motivated at least in part by a desire to avoid detection (Hossain et al. 2016, Meek et al. 2019, Dobbins et al. 2020).

Compared with the other methods described above, passive acoustic monitoring has the advantage that it does not require active participation on the part of hunters, nor does it require animals to be captured. A few studies have explored the use of passive acoustic monitoring to study hunting (both legal and illegal). Astaras et al. (2017) and Wrege et al. (2017) used acoustic monitoring to document the temporal patterns of hunting in Cameroon, Gabon, and the Republic of Congo. Dobbins et al. (2020) combined acoustic monitoring with cameras to assess spatial patterns of hunting in relation to prey densities in protected areas in Belize. Astaras et al. (2020) used acoustic monitoring to assess the effectiveness of increased anti-poacher patrols on hunting activity in a national park in Cameroon. Other studies have presented proof-of-concept results related to the capacity for passive acoustic monitoring to be used to detect gunshots (Hill et al. 2018, Singh et al. 2021, Wijers et al. 2021). Together, these studies suggest that the emergence of new acoustic monitoring technologies can be leveraged to understand and quantify hunting pressure with a resolution that was previously impossible, and that this understanding may translate into more efficient use of limited conservation resources (Astaras et al. 2020). Relative to other methods for studying hunting, acoustic monitoring remains a novel and underused tool.

Considerably more research has focused on detecting gunshots for policing and military applications. In many cases, seemingly accurate results have been obtained (Choi et al. 2014), but these often use proprietary techniques (e.g., ShotSpotter, Fremont, CA, USA) or complex signal processing analyses that will be unfamiliar or out of reach for most practicing biologists (Chacón-Rodríguez et al. 2011, Ahmed et al. 2013). Moreover, many published tests rely on recordings made under experimental conditions or generated artificially (Freire and Apolinario 2010, Chacón-Rodríguez et al. 2011). Whether these results generalize to the conditions typically encountered in natural areas, which include rain, wind, and a wide array of other biotic and abiotic noise sources, is largely unknown.

Significant questions remain to be answered with respect to the best practices for processing acoustic data to detect gunshots, the types of insights into hunting activity that can be derived from acoustic gunshot detections, and how acoustic monitoring compares with more established monitoring techniques. We examined the value of passive acoustic monitoring as a tool for detecting gunshots to study the spatiotemporal patterns of hunters. We had 3 primary aims. First, we sought to compare 2 methods for detecting gunshots in acoustic recordings using traditional accuracy metrics such as precision and recall and explored the implications of method selection for data quality. Second, we sought to examine the spatial and temporal patterns of hunting in our study area. We tested whether hunting rates were related to proximity to road access points, and whether hunting varied systematically across days of the week. Third, we sought to compare the detection rates of hunters using acoustic monitoring to the detection rates using camera trapping, a method that is more established for monitoring ecosystems and likely more familiar to biologists and conservation practitioners.

STUDY AREA

Our study took place in the Cooking Lake Blackfoot Provincial Recreation Area, which is a protected area in central Alberta, Canada (Figure 1; 53.49°N, 112.84°W). The field research ran from 1 September 2018 to 30 November 2018. Our study spanned the entire park, which measures up to 14.6 km along the east-west axis and 9.7 km along the north-south axis, for a total area of 9,700 ha. The 3 main landcover categories are treeless pastures, with cover dominated by grasses and other herbaceous vegetation; forests dominated by trembling aspen (*Populus tremuloides*) with lower numbers of balsam poplar (*P. balsamifera*), willows (*Salix* spp.), paper birch (*Betula papyrifera*), and white spruce (*Picea glauca*); and hundreds of wetlands and small lakes.

The park is managed by Alberta Parks as a multi-use site that accommodates a wide range of activities, including hiking, horseback riding, mountain biking, cross country skiing, and hunting, and industrial uses such as natural gas extraction and cattle grazing. Most licensed hunting occurs from September to November. Hunting in September and October focuses on waterfowl (family Anatidae), ruffed grouse (*Bonasa umbellus*), and archery hunting of big game. In November, big game firearm hunting occurs for white-tailed deer (*Odocoileus virginianus*), elk (*Cervus canadensis*), and moose (*Alces alces*). Motorized transport is prohibited in the park, so access occurs via walking, biking, or horseback, primarily along an expansive network of established trails (Figure 1).

Hunting is prohibited for all hunters in the park on Sundays, and guns can only be discharged during legal shooting light, which in Alberta runs from half an hour before sunrise until half an hour after sunset. Hereafter, we use the term daylight hours to refer to this period of time in which shooting is legal. Recreational or target shooting is prohibited in the park but permitted on adjacent private land.

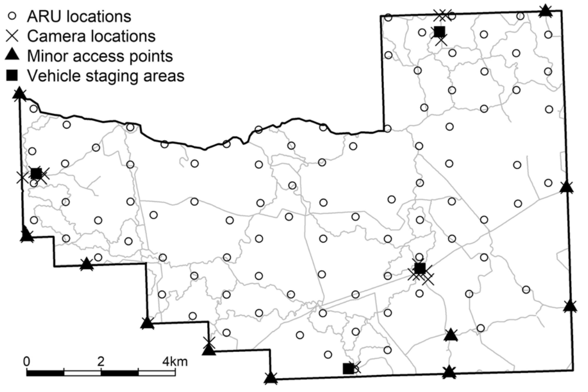

FIGURE 1 The study area where we placed autonomous recording units (ARUs) to estimate gunfire in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. Human access to the park occurs via minor access points (triangles) and vehicle staging areas (squares). Travel within the park is primarily by foot along trails and gravel roads (gray lines).

METHODS

Field deployments of recording units and remote cameras

We conducted a pilot study to determine the distance at which gunshots could be detected in our study area, and to determine the appropriate sample rate at which to record. Our goal was to achieve even spatial coverage of the study area while avoiding over-sampling and the collection of excess data. This required estimating 2 parameters prior to finalizing the study design: 1) the distance to which gunshots could be detected, to help us determine how far apart to separate microphones; and 2) the frequency content of gunshots, to help us determine the sampling rate with which to record. On 20 July 2018, we set up 15 autonomous recording units (ARUs, model: Song Meter SM4, Wildlife Acoustics, Inc., Maynard, MA, USA) at 100-m increments starting from a 200-m distance up to a maximum distance of 1,600 m in a forested environment within the study area. There was no precipitation at the time of the experiment, but winds registered Beaufort wind scale force 3 (average = 14.5 km/hr), with gusts reaching force 5 (~35 km/hr) according to the nearest weather station 15 km north of the park. We fired 4 types of guns that typified those used by hunters in the park, angling obliquely away from the ARUs: a .22-caliber long rifle (.22LR; 10 shots with 40-grain high velocity ammunition), a 12-gauge shotgun (10 shots, 5 each with 2.75 inch ammunition with a 28-g target load, and 3 inch ammunition with a 35-g magnum load), a .410 shotgun (5 shots with 2.5 inch ammunition with a 14-g load), and a centerfire rifle (10 shots, .308 Winchester 150-grain ammunition). Two technicians then listened to the recordings to assess whether each shot ($n = 30$) was audible on each ARU. All gunshots within 400 m were audible. Thereafter, the number of audible gunshots declined to 28, 24, 12, 9, and 5, as the distance increased from 500 m to 900 m. No gunshots were audible at or beyond 1,000 m within the test environment and in the conditions present that day. Based on this experiment, we decided to deploy microphones in a hexagonal grid, with adjacent microphones separated by 1,000 m.

We used the localpeaks function in the seewave package (Sueur et al. 2008) in R (R Core Team 2019) to calculate the peak frequency of each gunshot on the ARUs placed at the 200-m distance. Of 30 gunshots, the peak frequencies ranged from 370 Hz to 1,642 Hz, with notable differences between gun types (.22LR: $n = 5$, range = 1,000–1,642 Hz; shotgun: $n = 10$, range = 370–1,056 Hz; centerfire rifle: $n = 10$, range = 452–916 Hz). Because the lowest available sampling rate was 8,000 Hz in the SM4 units, we deemed that this recording setting would provide ample coverage of the frequencies relevant for detecting gunshots, while also minimizing data storage costs.

Based on the pilot study, we deployed SM4 ARUs at 91 locations across the study area according to a hexagonal grid with adjacent grid centroids separated by 1 km (Figure 1). We set all ARUs to record with an 8,000-Hz sample rate, 12.5 dB of gain, and to save recordings as 30-minute long .wav files. Prior to deployment, we calibrated each ARU to ensure consistency across units using a pistonphone (Sound Calibrator model 407744, Extech Instruments, Nashua, NH, USA) according to the protocol of the Bioacoustic Unit (Bioacoustic Unit 2019). This procedure ensured that all ARUs had similar responses to a 1-kHz tone broadcast at 94 dB. We replaced recording units whose response deviated by >4 dB from the average response across all units.

Our aim was to have all ARUs record continuously and simultaneously from 1 hour before sunrise until 1 hour after sunset, every other day from the beginning of September 2018 until the end of November 2018. However, unforeseen technical glitches led many recording units to deviate from the desired schedule. Specifically, some units recorded on odd days ($n = 28$), and others on even days ($n = 54$). These units followed the appropriate diurnal schedule (1 hr before sunrise to 1 hr after sunset). Still others ($n = 10$) recorded continuously, 24 hours per day until batteries failed (average = 18 days). When we switched batteries halfway through the project, many units switched from recording on even days to recording on odd days, or vice versa.

Although these issues were not desirable, they are unlikely to have affected our conclusions. Most importantly, we conducted our main analyses using rates (gunshots/hr) or included offsets, which effectively converts count data to rates. Still, to ensure that our recording effort was not markedly skewed with respect to the variables of interest, we analyzed per-station recording effort as a function of the main variables of interest, which were day of the week (one-way analysis of variance [ANOVA]), distance from access points (Pearson's correlation), and identity of the nearest staging area (one-way ANOVA).

In addition to the ARUs, we deployed remote motion-triggered cameras (PC800 Hyperfire Professional IR, Reconyx, Holmen, WI, USA) at 26 locations in the park. We placed cameras at trailheads at the 11 minor access points and 4 primary staging areas where people enter the park (Figure 1). We programmed cameras to take photos in bursts of 3 photos each time the motion sensor was triggered (saved as .jpg files). Five cameras sustained damage or did not function for the full season, so we restricted our analysis to data from the remaining 21 camera locations.

One observer manually processed the resulting 128,550 photos using the WildTrax remote camera processing platform (WildTrax, Edmonton, AB, Canada). The observer tagged and counted humans carrying firearms or bows as hunters, and tagged all other humans as non-hunters.

Following deployment and once recording had commenced, we conducted an additional experiment to assess the detectability of gunshots on ARUs in real-world conditions. We fired 19 gunshots from 2 types of guns (.22LR, $n = 10$; 12-gauge shotgun, $n = 9$) from 5 different locations on 17 and 18 October 2018. An observer manually inspected recordings in Audacity version 2.1.3 (Audacity Team 2017) to assess whether each gunshot was audible on ARUs within 1,600 m of the gunshot. To do so, the observer opened the relevant recording, searched for the gunshot blast around the time of interest, and noted whether it was audible or not.

Automated detection of gunshots in acoustic recordings

Sound level diminishes with distance from the source via attenuation. Thus, received sound levels are likely to contain spatial information relating to the distance of the gunshot from the microphone, which can provide important context during method evaluation and spatial analysis. We measured the relative sound level (RSL) of every gunshot detected by taking the maximum amplitude value among the spectrogram pixels between the start and end time of the detection. We calculated spectrograms using the spectro function (settings: window length = 512, overlap = 50%, not normalized) in the package seewave (Sueur et al. 2008) in R version 3.6.1 (R Core Team 2019). A spectrogram decomposes audio into short time and frequency bins, each with an associated amplitude value. We discarded frequencies below 200 Hz to reduce the influence of wind and other low-frequency noise, and frequencies above 1,500 Hz because gunshot energy was concentrated below this frequency in the pilot study, so that RSL values represented the maximum amplitude within the 200–1,500 Hz frequency range between the start and end time of the detection.

We evaluated 2 automated methods for detecting gunshots: an automated recognizer trained to detect the acoustic signatures of gunshots and an energy detector that detected all sounds exceeding a pre-specified RSL threshold. We compared the results from both techniques against results from a randomly selected subset of recordings from all 91 recording locations that a human observer manually scanned. Observers listened to gunshots with high-quality over-the-ear headphones (Sennheiser HD280 pro, Sennheiser Electronic, Wedemark, Germany, or Plantronics Blackwire 3220, Plantronics, Inc., Santa Cruz, CA, USA) to assess whether a sound was a gunshot or not. We used the freely available program Song Scope (Wildlife Acoustics) to train recognizers. Song Scope recognizers are based on hidden Markov models to detect patterns in acoustic signals. Song Scope assigns a score ranging from 0 to 100 to each detection indicating how well the sound fits the recognizer model. We trained the recognizer using the 30 gunshots fired during the field experiment. We then ran it through data from 4 of our recording locations, validated all of the results with scores >65, and extracted the true positive gunshot detections. We selected a score of 65 to limit the time needed for validation because the number of false positives increased exponentially at lower scores and became prohibitive to validate even at a modestly lower threshold of 60. Validating involved checking the spectrogram and, if needed, listening to the sound, to assess whether the detection was a gunshot or not. We then added these to the training set, to create a larger training set of 503 gunshots to train the final recognizer. We ran the final model (hereafter the recognizer) on the full acoustic dataset and 3 observers validated all detections that exceeded the score threshold of 65. We validated detections in Song Scope, which allows the user to view the spectrogram while listening to the sound.

The energy detector scanned for noises exceeding a pre-specified RSL threshold of 69 dB. Our goal in selecting the RSL threshold was to minimize the likelihood of double-counting a single gunshot at multiple microphones. We used the validated recognizer data for this purpose. First, we identified gunshots that were detected on multiple ARUs within 1 minute of each other and assigned them to distinct temporal clusters. The 1-minute allowance accounted for the fact that SM4 clocks drift out of synchrony over time. They have a specified accuracy of >3.5 parts per million (Wildlife Acoustics 2016) meaning that over the course of 3 months they are expected to drift by <30 seconds. The full recognizer dataset contained 1,774 distinct clusters of gunshot detections, with each cluster being detected on an average of 1.94 ARUs, implying that in the recognizer data, gunshots were frequently double-counted. We then iteratively filtered the dataset by removing gunshots measuring below a given RSL threshold, which had the effect of reducing double counting. We increased the threshold until every temporal cluster was detected on only one ARU, which occurred when the threshold was 69 dB.

To scan long recordings, we used the procedure for RSL measurement described above with a single difference: rather than specifying a start and end time, we scanned recordings in full to generate a time series of RSL within each recording. We defined detections as portions of recordings that exceeded 69 dB. We inspected detections in Praat (Boersma and Weenink 2014), which like Song Scope allows the user to view spectrograms and listen to the sound to confirm whether it was a gunshot or not.

We manually scanned a subset of recordings to assess the rates of false negatives produced by the 2 methods. We selected the subset of recordings by randomly selecting sound files from each recording location. The subset contained 391 recordings, with each of the 91 recording locations having either 4 or 5 recordings included. A technician viewed spectrograms of each recording in Audacity, and when they found a sound that resembled the pattern of a gunshot (i.e., having a sharp onset, broadband energy pattern, and long reverberations), they listened to it to confirm. To ensure consistency with earlier results, and because technicians expressed some uncertainty over whether some impulsive sounds (e.g., machinery, shunting trains, snapping branches) were gunshots or not, one observer (RWH) checked the resulting detections and removed those that were deemed not to be gunshots (n = 555 dubious noises removed). The final dataset contained 639 gunshots.

Comparison of gunshot detection methods

We compared the 2 automated methods for gunshot detection against the subset of data processed by manual scanning. When comparing a method against the manual scanning dataset, the parameter of interest was recall—the proportion of total gunshots detected by the automated technique. However, we knew *a priori* that the energy detector would not detect gunshots <69 dB, and the manual scanning dataset contained no gunshots ≥69 dB. Therefore, we also compared the 2 automated methods against each other using the full set of recordings. In this comparison, we restricted the comparison to sounds registering ≥69 dB, which allowed for a more informative comparison of the performance of the energy detector.

Data processing for the automated methods was a 2-step process: automated detection followed by manual validation to remove false positives. We recognized that the validation step could introduce errors in which a true gunshot was flagged as a false positive or a non-gunshot was flagged as a true positive. We also hypothesized that the rates of such errors might differ between the 2 automated methods because the recognizer detected many faint sounds and the energy detector only detected loud sounds >69 dB RSL. Although it was not possible to identify validation errors directly, we made a close approximation by assessing the rates of inter-observer disagreements during manual processing steps. Logically, if a sound was flagged as a negative in one dataset but flagged as a positive in another, one observer must have been wrong, even if it is uncertain which one. A dataset producing many ambiguous detections can be considered lower quality than one containing few because it is less robust to scrutiny and less reproducible.

For the recognizer, we assessed the rate at which sounds were initially detected by the recognizer but validated as false positives, while being classified as true positives in the manual scanning dataset. We also assessed the converse: the rate at which sounds were passed over (i.e., classified as negatives) in the manual scanning dataset, while being detected and validated as true positives in the recognizer data. To make a comparable assessment for the energy detector, we assessed the rate at which detections >69 dB RSL were validated as a true positive in the recognizer dataset and false positive in the energy detector dataset and vice versa.

Modeling of spatial and temporal patterns of gunshots

We used a generalized linear mixed model (GLMM) to assess spatial and temporal patterns of hunting in the park (Zuur et al. 2009). Because of the poor recall and high levels of inter-observer disagreements in the recognizer data, we used the energy detector data for this and subsequent analyses. We converted the data from each station to a daily detection history, which included the number of gunshots detected on that day at that station, the number of daylight hours during which the ARU was active, and the day of the week. When consecutive gunshots were spaced by <1 minute from the previous gunshot, we assumed these came from the same hunter and counted them as a single event. In some cases, very large numbers of gunshots (up to a few dozen) were detected in a short period of time, occurring in bursts of 3 to 5 shots. These patterns were always detected on ARUs near the edge of the park, especially at a few ARUs along the southern edge where there are shooting ranges on nearby private land. Waterfowl hunting may present similar short volleys of shots, but these target shooting events could be identified by their occurrence within longer periods of sustained shooting atypical of hunting and by their occurrence at predictable locations. Although we could not be certain, we assumed these shots resulted from target shooting outside the park and removed them from the data set.

We modeled the response (number of gunshots detected on each station-day) using a negative binomial GLMM with a log link, with an offset for the number of hours the ARU was active on each day during daylight hours. The independent variables were 1) the distance from the ARU to the nearest staging area or minor access point (continuous variable), 2) the day of the week (categorical variable with 7 levels), and 3) the identity of the nearest staging area (categorical variable with 4 levels). This latter variable had the effect of dividing the park into 4 roughly equivalent regions (west, southwest, southeast, and northeast; Figure 1), which allowed an assessment of broad spatial hunting patterns. We included the distance variable as either a linear or quadratic term. To account for multiple days of recording from each location, we also included location in all models as a random intercept, which we assumed was normally distributed with a mean 0 and variance σ^2. We compared 12 models that included all possible combinations of the fixed effects and selected the best model as the model with the lowest value of Akaike's Information Criterion value corrected for small sample sizes (AIC_c; Burnham and Anderson 2002). We also calculated AIC_c weights to determine if any models other than the top model should be given further consideration. We carried out this analysis using the lme4 package (Bates et al. 2015) in R. We used the functions provided in the DHARMa package (Hartig 2021) to check model fit and to confirm that the model was not over-dispersed. Equivalent models fit using a Poisson GLMM showed evidence of over-dispersion, which supported our use of the negative binomial GLMM.

Comparison of ARU data and remote camera data

To assess whether ARUs and cameras produced similar spatial and temporal patterns of hunter activity, we directly compared the data from each. For the temporal comparison, we calculated the rate of gunshots detected via the energy detector on the different weekdays. To do so, we divided the number of gunshots detected on each weekday by the total hours of recording during daylight hours. We multiplied this value by 10 hours, which was the mean number of daylight hours per day during the study period, to estimate the average number of gunshots detected per ARU for each weekday. For the cameras, we simply calculated the number of hunters detected per camera day for each weekday during the study period.

The spatial comparison was less straightforward because the ARUs and cameras were not deployed in the exact same locations. Therefore, we assigned each ARU and camera location to the nearest access point (staging area or minor access point; Figure 1). We summarized ARU data as gunshots per daylight recording hour at the ARUs near each access point. As above, we multiplied this value by 10 to convert the rate to gunshot detections per day. We summarized the cameras by dividing the number of hunters photographed by the total number of camera-days among the cameras nearest to each access point. Three of the 15 access points lacked data for cameras or ARUs, so we made the comparison among the remaining 12 access points. We compared the weekday ($n = 7$ weekdays) and spatial patterns ($n = 12$ access points) of detections between ARUs and cameras using the cor.test function in R.

RESULTS

Field recordings and gunshots from known locations

Our dataset comprised 42,420 hours of recording from 91 locations across Cooking Lake Blackfoot Provincial Recreation Area. Despite the technological glitches described above, recording effort was well distributed across the variables of interest (Figure 2). Minor differences in per-station recording effort occurred across days of the week (Figure 2A; ANOVA: $F_{6,630} = 3.217$, $P = 0.004$). Data did not indicate that recording effort was related either positively or negatively to the distance from access points (Figure 2B; $r_{89} = -0.168$, $P = 0.11$). Average per-station recording effort was only slightly different among staging areas (Figure 2C; ANOVA: $F_{3,87} = 3.80$, $P = 0.013$). Because our analyses relied on the conversion of counts of gunshots into rates of gunshots per hour, these slight recording effort differences were unlikely to have affected our conclusions.

Data from guns fired at known locations during the study confirmed that the relative sound level of gunshots declined with increasing distance from an ARU (Figure 3; $r_{52} = -0.50$, $P < 0.001$). Among the shots from known locations, 7 shotgun blasts exceeded the 69 dB sound level threshold specified for the energy detector but no

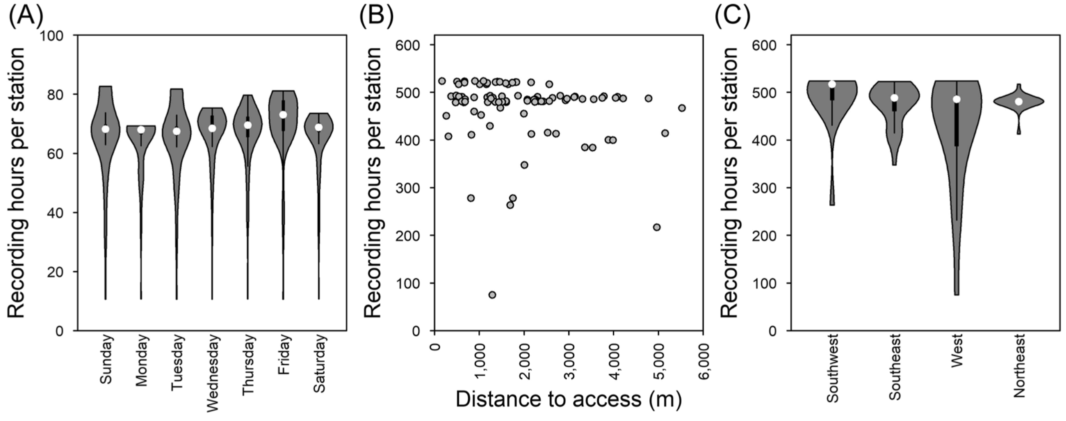

FIGURE 2 Visual representation of the relationship between recording effort and 3 variables of interest for autonomous recording units used to estimate gunfire in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. A) Recording hours per station differed slightly across days of the week. Total recording effort varied from 5,803 hours on Mondays to 6,424 hours on Fridays. B) Recording hours per station was not systematically related to the distance from the nearest access point. C) Average recording hours per station differed with slightly more hours in the Southwest than elsewhere.

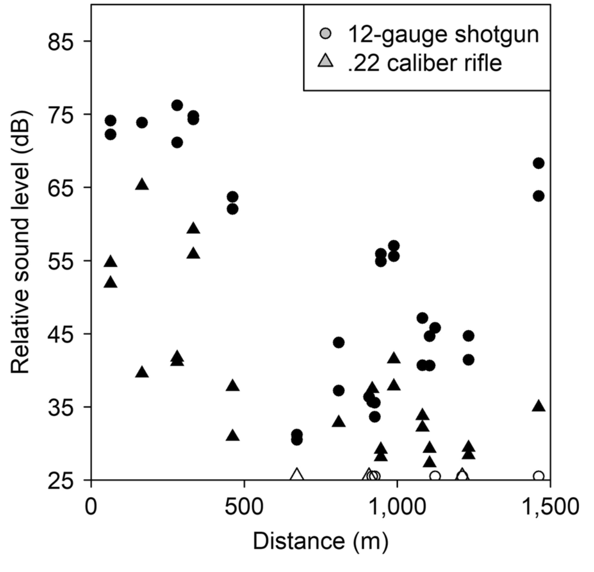

FIGURE 3 Relative sound level (RSL) of gunshots fired from known locations was strongly related to the distance of the gunshot from the autonomous recording unit (ARU) in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. Black shapes indicate shots that were audible on the ARU and for which we measured RSL. White shapes were not audible on the ARU likely because they were below the noise floor of the recording; we did not make RSL measurements for these shots, but for visualization purposes we show them along the bottom edge of the plot.

.22-caliber rifle blasts did, which implied the .22-caliber rifles produced quieter blasts. All shotgun blasts that exceeded the 69 dB threshold were <334 m from the ARU. Some gunshots were inaudible as close as 672 m from the ARU, while others were audible to 1,461 m. We consider it very likely that some gunshots would be audible well beyond 1,600 m in certain conditions (Figure 3), though we did not check recordings beyond this distance.

Comparison of gunshot detection methods

The Song Scope recognizer produced 149,980 detections that exceeded the specified score threshold of 65. After validation, 7,663 were classified as gunshots, for a precision of 0.051. The RSL of the resulting gunshots were 43.6 ± 9.8 dB ($\bar{x}$ ± SD). Although we did not systematically track the sources of false positives, it was our impression that most were caused by sounds such as wind, rain, machinery, beaver (*Castor canadensis*) tail slaps, and other impulsive sounds. The manual scanning method produced 639 gunshots in 202 recording hours. Of these, 21 were detected in the validated recognizer results, giving the recognizer an estimated recall of 0.033.

We found high levels of inter-observer disagreements when comparing the recognizer data to the manual scanning data (Table 1). Twenty sounds were classified as true positives during recognizer validation and as negatives during manual scanning, and a further 16 sounds were classified as false positives during recognizer validation and as true positives during manual scanning. Therefore, of 57 sounds that were reviewed by both observers and classified as gunshots by at least one observer, the second observer agreed it was a gunshot just 37% of the time (21 out of 57). Instances of observer disagreement were associated with lower RSLs (39.7 ± 4.1 dB) than instances of observer agreement (43.3 ± 5.6 dB; $t_{32.3}$ = 2.6, P = 0.015), suggesting that inter-observer disagreements may have been caused at least in part by faintness.

The energy detector identified 39,701 events that exceeded the 69 dB threshold, of which 563 were validated as gunshots, for a precision of 0.014. The RSL of the sounds were 72.0 ± 2.5 dB, so they were about 28.4 dB louder than the sounds detected by the recognizer. The energy detector identified none of the 639 gunshots in the manual scanning dataset (recall ≈ 0) because none exceeded the 69 dB threshold (range = 9.0–67.3 dB). Therefore, we instead assessed recall by comparison with the validated recognizer data. Because the recall of the recognizer was 0.033, and the energy detector detected 93% fewer total gunshots, the recall of the energy detector can be assumed to be about 0.0023 on the full dataset. However, this assessment includes gunshots with any RSL value,

TABLE 1 Comparison of gunshot detections from recognizer data from autonomous recording units (ARUs), manual scanning, and energy detector data from ARUs, while separating the automated detection process from the observer-based validation process, in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. Detection (d) and validation (v) could not be separated in the manual scanning data because a human observer conducted both processes simultaneously. We compared data from the recognizer and manual scanning for the 391 recordings that were manually scanned and compared data from the recognizer and energy detector on events exceeding 69 dB. We separate data by sounds detected by the automated process and labeled a true positive by an observer (d+,v+), sounds detected by the automated process and labeled a false positive by an observer (d+,v−), sounds labeled as positives (v+) or negatives (v−) during the manual scan, and sounds not detected, and thus not subjected to validation (d−). Sounds detected by neither method could not be counted as discrete events (N/A) but represent the rest of the dataset.

	Recognizer		
	d+ , v+	**d+ , v−**	**d−**
Manual scan			
v+	21	16[a]	602
v−	20[a]	488	N/A
Energy detector			
d+, v+	78	1[a]	484
d+, v−	0[a]	338	38,800
d−	0	0	N/A

[a]Observer disagreements in which the human observers differed with respect to whether a sound was a gunshot or not.

whereas only gunshots >69 dB RSL are detectable by the energy detector. The validated recognizer data contained 78 gunshots exceeding 69 dB RSL, and among these 100% were detected by the energy detector and validated as true positives (recall ≈ 1 on sounds ≥69 dB). Recall of the energy detector was therefore either very low (near zero) or very high (near one), depending on whether recall was assessed by comparison with all audible gunshots or only loud gunshots.

The energy detector generated data with much lower rates of inter-observer disagreements (Table 1). Out of 79 events ≥69 dB that were reviewed by observers validating the recognizer data and the energy detector and determined to be gunshots by at least one observer, there was agreement on 78 of them (98.7%). Energy detector data could thus be considered high quality, whereas the low recall and low levels of observer agreements in the recognizer data led us to question its validity in some cases. We used the energy detector data for all subsequent analyses of the spatiotemporal patterns of hunting.

Modeling of spatial and temporal patterns of gunshots

Modeling was based on 3,874 station-days in which 249 gunshot events were detected with the energy detector. The top model selected via AIC_c was the full model including all predictor variables (Table 2). The second model was very similar to the top model but included a linear, rather than quadratic, effect of distance from the nearest access point; together these 2 top models were ascribed 99% of the model weight. Gunshot rates declined as a function of an ARU's distance from an access point (Figure 4A; Table 3). Among ARUs placed farther than 3 km from access points, the average rate of gunshot detections was 1.05 gunshot per 1,000 recording hours. In contrast, ARUs placed within 3 km of an access point detected gunshots at a rate of 7.0 gunshots per 1,000 hours. There were also

TABLE 2 Models predicting gunshot detections by the energy detector data from autonomous recording units (ARUs) in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. Variables included quadratic ($Dist^2$) and linear (Dist) effects of the distance of an ARU from the nearest access point, day of the week (day) as a categorical variable with 7 levels (Sunday–Saturday), and the identity of the nearest of the 4 major staging areas (area) as a categorical variable with 4 levels (west, southwest, southeast, and northeast). For each model, we present the number of parameters (*K*), Akaike's Information Criterion corrected for small sample size (AIC_c), difference in AIC_c (ΔAIC_c), model weight, and log-likelihood.

Variables	*K*	AIC_c	ΔAIC_c	Weight	Log-likelihood
$Dist^2$ + day + area	14	1,668.15	0.00	0.60	−820.02
Dist + day + area	13	1,669.00	0.85	0.39	−821.46
$Dist^2$ + day	11	1,677.51	9.36	0.01	−827.72
Dist + day	10	1,678.02	9.87	0.00	−828.98
Day + area	12	1,686.89	18.74	0.00	−831.4
Day	9	1,697.59	29.44	0.00	−839.77
$Dist^2$ + area	8	1,741.96	73.81	0.00	−862.96
Dist + area	7	1,742.92	74.77	0.00	−864.45
$Dist^2$	5	1,752.51	84.36	0.00	−871.25
Dist	4	1,752.98	84.83	0.00	−872.49
Area	6	1,759.74	91.59	0.00	−873.86
Null	3	1,771.36	103.21	0.00	−882.68

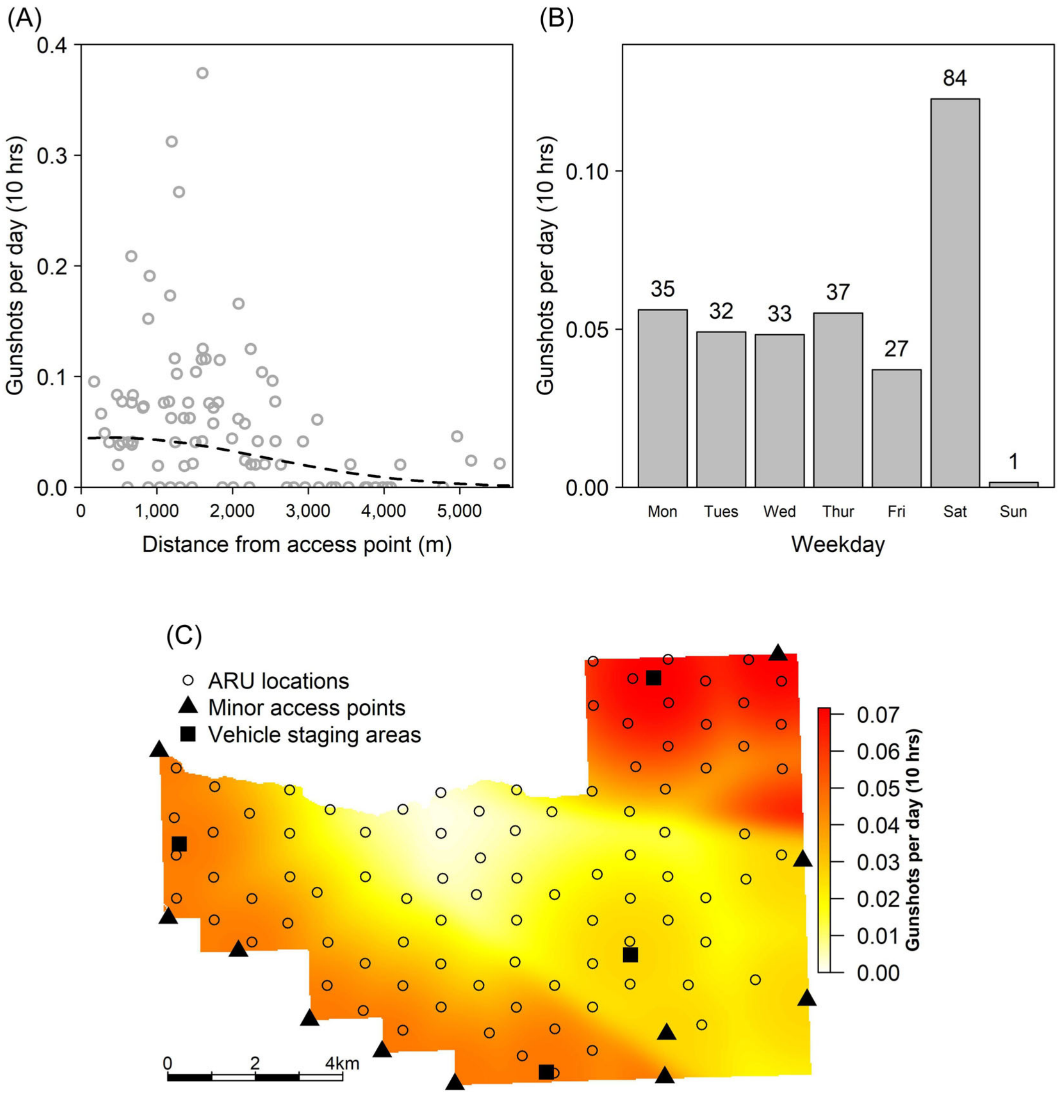

FIGURE 4 The rates of gunshots detected on autonomous recording units (ARUs) varied significantly across space and time in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. A) Gunshots were detected at high rates near access points, and low rates in the remote interior of the park. Gray circles indicate station averages, and the black line indicates the model predictions with all other covariates held at their average value. B) Gunshots were detected at high rates on Saturdays, moderate rates on weekdays, and very low rates on Sunday. Numbers above bars indicate the total number of gunshots detected on each day of the week. C) Predicted hunting intensity varied significantly across the park, with highest intensity in the northeast, and lowest intensity in the center of the park. Values depicted on the raster are predicted hunting rates for every location in the park, averaged across the 7 weekdays (30-m spatial resolution).

regional differences in hunting intensity within the park with the highest intensity of hunting in the northeastern portion of the park (Figure 4C; Table 3).

Rates of gunshot detections also varied as a function of the day of the week, with our model determining that gunshots were detected at roughly twice the rate on Saturday as from Monday to Friday (Figure 4B; Table 3). Only one shot was detected on a Sunday, when hunting in the park is prohibited, in 6,086 recording hours.

TABLE 3 Coefficient estimates for the best supported model (the model with all covariates) for describing the spatiotemporal patterns of hunting activity as measured by gunfire detected by autonomous recording units (ARUs) in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. The estimated value for random effect of ARU location ($\sigma_{Location}$) is 0.47.

Variable	Estimate	SE	P
Intercept	−5.31	0.26	<0.001
Distance[a]			
Distance	−38.16	8.90	<0.001
Distance^2	−13.47	8.09	0.1
Day[b]			
Monday	0.41	0.29	0.15
Tuesday	0.28	0.29	0.34
Wednesday	0.26	0.29	0.37
Thursday	0.39	0.28	0.16
Saturday	1.19	0.26	<0.001
Sunday	−3.22	1.03	0.002
Nearest staging area[c]			
Southeast	−1.01	0.25	<0.001
Southwest	−0.43	0.27	0.11
West	−0.46	0.25	0.07

[a]Distances were scaled to have mean = 0 and SD = 1 prior to modeling.
[b]Day coefficients were estimated relative to a Friday reference.
[c]Nearest staging area coefficients were estimated relative to the northeast staging area.

Comparison of ARU data and remote camera data

The remote cameras detected 2,417 hunter entries or exits at 21 locations. The rate of detection of hunters showed a nearly identical weekly pattern to the temporal pattern of gunshot detections (Figure 5A; $r_5 = 0.94$, $P = 0.001$). As in the ARU dataset, the rate of hunter detections was about twice as high on Saturday as on weekdays, and hunters were nearly absent on Sunday. The most notable difference between the 2 datasets was that the daily rate of hunter detections on cameras was consistently much higher than the rate of gunshot detections on ARUs (Figure 5A). Collapsing the 2 datasets to the nearest access point produced detection rate data from both cameras and ARUs for 12 access points. Access points with more hunter detections on cameras also tended to have higher rates of gunshots on nearby ARUs (Figure 5B; $r_{10} = 0.62$, $P = 0.030$).

DISCUSSION

Comparison of gunshot detection methods

We collected a large dataset of over 40,000 hours of acoustic recordings to assess the utility of ARUs for monitoring hunting-related gunshot activity. Our results indicate that the data derived from ARUs could be used to infer spatial patterns of hunting that broadly matched expectations: rates of gunshots declined with increasing

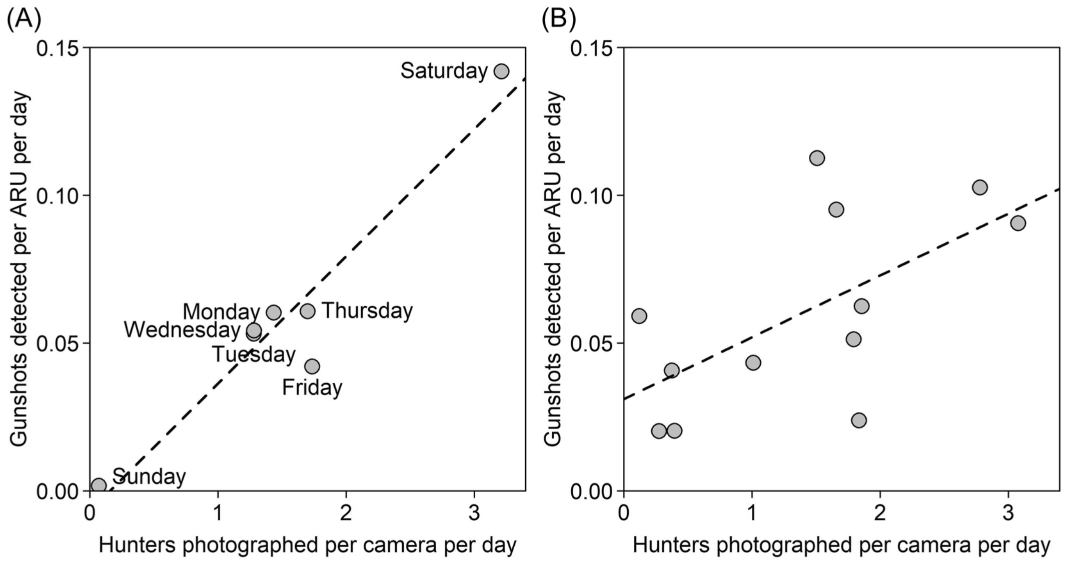

FIGURE 5 Comparison of the rates of gunshots detected on autonomous recording units (ARUs) and hunters detected via remote cameras in the Cooking Lake Blackfoot Provincial Recreation Area in central Alberta, Canada, in 2018. A) Both methods detected similar temporal patterns of hunting activities with the highest rates of hunting activity on Saturdays, moderate rates on weekdays, and near-zero rates on Sundays when hunting is prohibited. B) Spatial patterns in the rates of detections near different park access points indicate the 2 methods produced rates of detections that were positively correlated. Points show average rates of detections for the ARU (*y*-axis) or camera (*x*-axis) locations nearest to 12 different access points into the park.

distance from access points and gunshots were seldom detected in the center of the park. We also found high variation in the rates of gunshots across the 7 days of the week, which matched expectations. Rates of gunshots were highest on Saturdays when hunters likely had days off from work, rates were at intermediate levels on weekdays, and gunshots were almost absent on Sundays when hunting in the park is prohibited. Our results highlight the potential of passive acoustics as a tool for monitoring of hunting activities and enforcement of regulations. Further, the global growth of acoustic monitoring programs focused on biodiversity (e.g., Sugai and Llusia 2019) can serve a complementary purpose for monitoring recreational, industrial, or other human activities that produce sound.

One challenge facing any acoustic monitoring project is the problem of detecting events of interest from long duration recordings (Gibb et al. 2019). We compared 2 different approaches for detecting gunshots: a recognizer that was trained to detect the specific spectrotemporal characteristics of gunshots, and an energy detector that detected any sound with high relative sound levels, regardless of its finer acoustic features. Our eventual decision to use the much simpler energy detector rather than the recognizer for our final analysis warrants further discussion. Many studies using machine learning to analyze acoustic data place strong emphasis on a few common metrics of accuracy for comparing methods. The stated aim is to reduce false positives and false negatives relative to true positives and true negatives, and this is typically accomplished by collapsing these numbers into metrics such as precision and recall (Digby et al. 2013, Heinicke et al. 2015), or composite metrics such as area-under-the-curve and mean average precision (Joly et al. 2019, Stowell et al. 2019, Brooker et al. 2020). In our study, however, comparison of the 2 methods on the full dataset showed that both precision and recall were higher for the recognizer (0.051 and 0.033, respectively) than for the energy detector (0.014 and about 0.0024, respectively). In other words, a higher proportion of the detections made by the recognizer were true positives, and more of the gunshots in the full dataset were detected by the recognizer (7,663 gunshots for the recognizer vs. 563 gunshots for the energy detector).

Had we used precision and recall to invoke a simple decision rule, we would have selected the recognizer over the energy detector. Doing so, however, would have ignored what is perhaps the issue of most importance: data quality. It is inarguable that acquiring a high-quality dataset with minimal bias with respect to the parameters of interest is the ultimate goal; although, in our view this is often glossed over when performance assessments focus too heavily on accuracy metrics.

Despite its higher precision and higher recall overall, the recognizer fell short in 3 main ways. First, although the recognizer detected more total gunshots, these gunshots occurred across a wide range of RSL values and were, on average, nearly 30 dB fainter than those detected by the energy detector. Worryingly, at lower RSL values we found that when one observer detected a gunshot, the second observer seldom agreed it was a gunshot. It was not generally possible in our study to ascertain whether these errors were false negatives by one observer or false positives by the other—both of which have been shown to increase with distance and decline with signal-to-noise ratio in auditory surveys (Alldredge et al. 2007, Simons et al. 2007, McClintock et al. 2010*a*). Regardless of the type of error involved, we were much more comfortable with the high levels of inter-observer agreement (98.7%) in the energy detector data, which can be attributed to the absence of faint and ambiguous sounds.

Second, the recognizer missed most clear gunshots, including 86% of the gunshots detected by the energy detector. Errors of this type may not be problematic if the errors occurred randomly, but they could introduce bias if the errors were non-random (McClintock et al. 2010*b*). In our case, we worried the recognizer might be biased towards detecting gunshots under certain weather or vegetation conditions, or toward detecting certain types of firearms over others; evaluating these possibilities would have required a significant investment of time and was well beyond the scope of this study. Missed detections by the energy detector were comparably easy to understand: if they were missed, it was because they were not loud enough to be detected (which generally meant the hunter was too far from the microphone; Figure 3). Our results still require some assumption that the received RSL was not strongly correlated with our model parameters (distance to access points, day of the week, and nearest staging area), but we considered these assumptions reasonable.

Third, by detecting gunshots with both high and low RSL, the recognizer often detected gunshots at great distance (Figure 3; Piña-Covarrubias et al. 2019, Dobbins et al. 2020). We suspect that the recognizer likely detected substantial numbers of gunshots that originated from outside of our study area or from distant areas of the park. Because a primary goal of our study was to study spatial patterns in hunter activity, this disconnect between the locations of gunshots and microphones may generate illusory spatial patterns of hunting activity, and therefore represented another limitation of the recognizer data. A related problem is that distant gunshots are more likely to be double-counted on multiple microphones.

Blanco et al. (2012) argued that as computational resources have facilitated the use of ever-more-sophisticated models, there has been an apparent decline in the attention paid by ecologists to data quality. Passive acoustic monitoring studies may face similar risks if too much emphasis is placed on simply accuracy metrics, with too little attention paid to the goal of producing high-quality, easily interpretable, and defensible data. Our study illustrated that the concerns faced during real-world applications extend well beyond composite measures of algorithm accuracy. We suspect that 2 results will generalize beyond this study. First, the performance of data processing methods, including the ranked performance of different techniques, might differ as a function of the relative sound level of the target sound (Digby et al. 2013). In our case, the ranking of the recognizer versus energy detector in terms of recall was entirely dependent on whether they were assessed on all audible sounds or only those exceeding the RSL threshold. Second, the human benchmark will never be perfectly accurate, and inter-observer agreement is likely to suffer as relative sound level declines. Both of these results have important implications for acoustic data processing, and most importantly, for ensuring the resultant data is of the highest possible quality.

Comparison of ARUs and cameras for detecting hunting activity

Autonomous recording units and cameras captured similar spatial and temporal patterns of hunting activity in our study area. Both methods showed peak hunting activity on Saturdays, very low activity on Sundays, and moderate activity on weekdays, and there was a positive correlation between the ARU and camera data in terms of the broad spatial patterns of activity in the park (Figure 5). Although the similarities between the 2 datasets imply that one monitoring technique might readily substitute for the other, the differences are equally important, and suggest that acoustic monitoring can be used to fill information gaps not easily filled by cameras (Buxton et al. 2018, Garland et al. 2020). When used as they were in this project, cameras at trail heads and staging areas can reveal the frequency and intensity of hunting (and other recreational access), and can show use type, such as travel mode, activity type, and gear type. The ARUs offer the ability to measure aggregated temporal and spatial patterns in the use of certain audible gear, such as firearms (or motorized recreation), that may be useful for managers tasked with ensuring hunter or visitor experiences in expansive areas.

The first obvious difference between ARUs and cameras is in the area surveyed. Depending on whether data are truncated via RSL or not, ARUs might survey an area from several hectares in size to several square kilometers (Figure 3). By contrast, a remote camera can only survey a few square meters. Cameras are expected to perform worse in open environments where hunters may not follow established trails, or when surveying poachers (Hossain et al. 2016, Astaras et al. 2017). For instance, in a direct comparison of cameras versus ARUs, Dobbins et al. (2020) found that ARUs detected almost 10 times as many hunters as did cameras. Our study, conversely, found much higher detection rates on cameras than ARUs, with the likely reason for the difference being that we deployed cameras to photograph legal hunters entering the study area via established trails, while Dobbins et al. (2020) studied illegal hunters who appeared intent on evading detection, including sometimes stealing or destroying cameras. The fact that ARUs detect hunting at such great distances suggests they are well-suited to monitoring poaching, and might provide a more comprehensive picture of hunting activity in areas where the precise routes taken by hunters are unpredictable. The second difference between ARUs and cameras is in the type of events they detect. Notably, the correlation between the 2 datasets does not mean the 2 methods measured the same parameter. Instead, the correlation likely resulted from a mechanistic link in which the rate of gunshots detected on ARUs was a function of the hunting effort on any given day as measured by cameras, consistent with previous findings that the amount of time spent hunting is a primary determinant of hunting success (Bhandari et al. 2006).

Given that ARUs can measure hunting outcomes defined as shots fired, and cameras can measure hunting effort in the absence of shooting, the ratio between the 2 may prove worthwhile as a proxy for hunter efficiency (shots fired at an animal per unit time or the inverse, the time needed to fire a shot at an animal). Theoretical models predict that spatial and temporal variation in hunter efficiency should strongly influence where and when hunters allocate their hunting effort (Van Deelen and Etter 2003), in which case methods that can directly estimate efficiency could be a useful management tool. Temporal changes in efficiency in an area may also closely track changes in animal density (Holsworth 1973). Perceptions around efficiency are also considered important in economic models of hunting activity, which suggest that the costs and benefits of where to hunt might be positively related to the perceived likelihood of successful harvests (i.e., predicted efficiency), and negatively related to factors such as travel costs (Balkan and Kahn 1988). Projects employing both ARUs and cameras may provide useful insights into the factors driving spatial variation in hunting effort, with important downstream consequences for wildlife managers.

Spatial and temporal patterns of hunting activity

Our results show high degrees of spatiotemporal variation in hunting activity. The temporal variation observed in this study provide an interesting contrast to the results described by Astaras et al. (2017) in a study in Korup

National Park in Cameroon. In that study, hunters focused their hunting effort mid-week, and also hunted largely at night. Although we did not record at night, we are confident that night-time hunting would be exceptionally rare in our study area, given that it is illegal and the hunters in our study appeared to generally follow regulations. The differences between the patterns seen in Korup National Park and our study can be attributed to the different behaviors exhibited by legal hunters and poachers. In our study, hunters had no reason to evade detection, and thus appeared to disproportionately hunt when it was convenient (i.e., on Saturdays) and avoid hunting when it was illegal (i.e., on Sundays). Poachers in other areas of the world, on the other hand, appear to conceal their activities by hunting when they are least likely to be caught by patrols.

Poaching is not a major problem in our study area. A greater concern is the potential for conflict between hunters and other park users. Our study area is just 40 km east of Edmonton, a city of over a million people that is expected to double in population in the next 50 years. If trends continue, use of the Cooking Lake-Blackfoot Provincial Recreation Area by hunters and non-hunters is expected to increase, with potential increases in conflicts surrounding the perceived risks to non-hunters.

Our results provide some of the information needed to ease these conflicts. Hunting was highest on Saturdays, which suggests that recreational users seeking to avoid hunters should be encouraged to visit on Sundays during hunting season, when hunting was almost completely absent. The absence of hunters on Sundays also points to high levels of compliance among hunters, which should be applauded and can also be used to counter negative stereotypes of hunters. Our examination of spatial patterns of hunting activity could similarly be used to encourage non-hunters to avoid the northeast area of the park during hunting season (Figure 4C). Non-hunters could also be advised to recreate in the center of the park, although we note that non-hunters face similar issues as hunters when it comes to accessing the remotest areas of the park.

Future directions

Our results highlight the potential of passive acoustic monitoring as an emerging tool for detecting and studying hunting activity. We detected significant spatial variation in hunting activity, including a clear pattern of hunters concentrating their effort near vehicle access points. This pattern is consistent with existing literature highlighting the important role of roads and vehicle access as one of the main predictors of hunting pressure (Fuller 1990, Ziegler et al. 2016). Compared with previously employed methods such as questionnaires or interviews (Johnson et al. 2016, Ziegler et al. 2016), ARUs reveal hunting patterns with very high spatial resolution—in our study, within a few hundred meters of the hunting event (Figure 3). Increasingly, ARUs are becoming commercially available that provide even finer resolution. When equipped with GPS units, the clocks of ARUs can be synchronized within about 1 ms (Mennill et al. 2012), which in a project such as ours would confer 2 main benefits. First, it would allow a much more accurate assessment of double-counted gunshots. We sometimes had difficulty ascertaining whether a gunshot detected on 2 microphones was a single gunshot or 2 different shots because of clock drift. Second, synchronized clocks allow acoustic events to be localized within a few meters of their true location (Wijers et al. 2021). Accurate gunshot localization will open up a suite of new research opportunities, such as allowing fine-scale assessments of the rates of hunting in different types of landscapes. It will also more concretely allow researchers to assess whether gunshots originated inside or outside a protected area. In this study, we strongly suspected that many or even most of the gunshots audible within the park originated from outside, which may lead visitors to over-estimate how much shooting actually occurs in the park. The better spatial resolution afforded by localization would help resolve this even further.

Finally, real-time monitoring might become feasible in the future, in which case acoustic monitoring might see increased use as an enforcement tool, rather than strictly for research. We suspect that a fully automated real-time system would require a vastly improved automated gunshot recognizer. At present, false positives are likely too numerous for real-time analysis to be practical (Wrege et al. 2017, Cardoso 2019). Emerging classification techniques, such as convolutional neural networks, may help considerably (Singh et al. 2021). With adequate research attention, and continued improvements in machine learning, real-time analysis seems like a logical and attainable long-term goal.

ACKNOWLEDGMENTS

We thank D. Vadnais, K. Musgrave, K. Henderson Pekarik, M. Dieleman, and M. Martel for assistance processing data. We thank M. Knaggs, L. Garland, J. Johnson, J. Kennedy, L. Leston, C. Leven and Government of Alberta staff, especially D. Vujnovic and K. Hayduk, for help with field work, and Juan Andrés Martínez-Lanfranco for statistical advice. Thanks to S. Brainerd (Associate Editor), A. Knipps (Editorial Assistant), A. Tunstall (Copy Editor) and A. Cox (Content Editor) for their review and suggestions that improved the manuscript. Salary support for R. W. Hedley was funded in part by an NSERC Postdoctoral Fellowship and a Mitacs Accelerate fellowship. Additional funding support was provided by an NSERC Discovery Grant to E. M. Bayne.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

No ethical information provided.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Figshare at https://doi.org/10.6084/m9.figshare.19397894.

ORCID

Richard W. Hedley http://orcid.org/0000-0001-9588-5515

REFERENCES

Ahmed, T., M. Uppal, and A. Muhammad. 2013. Improving efficiency and reliability of gunshot detection systems. IEEE International Conference on Acoustics, Speech and Signal Processing 2013:513–517.

Alberta Government. 2021. 2021 Alberta guide to hunting regulations. <https://open.alberta.ca/publications/1485-4287>. Accessed 25 Nov 2021.

Alldredge, M. W., T. R. Simons, and K. H. Pollock. 2007. Factors affecting aural detections of songbirds. Ecological Applications 17:948–955.

Astaras, C., J. M. Linder, P. Wrege, R. Orume, P. J. Johnson, and D. W. MacDonald. 2020. Boots on the ground: the role of passive acoustic monitoring in evaluating anti-poaching patrols. Environmental Conservation 47:213–216.

Astaras, C., J. M. Linder, P. Wrege, R. D. Orume, and D. W. Macdonald. 2017. Passive acoustic monitoring as a law enforcement tool for Afrotropical rainforests. Frontiers in Ecology and the Environment 15:233–234.

Audacity Team. 2017. Audacity: free audio editor and recorder. <https://audacityteam.org>. Accessed 20 May 2022.

Balkan, E., and J. R. Kahn. 1988. The value of changes in deer hunting quality: a travel cost approach. Applied Economics 20: 533–539.

Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

Beirne, C., A. C. Meier, A. E. Mbele, G. Menie Menie, G. Froese, J. Okouyi, and J. R. Poulsen. 2019. Participatory monitoring reveals village-centered gradients of mammalian defaunation in central Africa. Biological Conservation 233:228–238.

Bhandari, P., R. C. Stedman, A. E. Luloff, J. C. Finley, and D. R. Diefenbach. 2006. Effort versus motivation: factors affecting antlered and antlerless deer harvest success in Pennsylvania. Human Dimensions of Wildlife 11:423–436.

Bioacoustic Unit. 2019. SongMeter (SM4) Maintenance Protocol. <https://www.wildtrax.ca/home/resources/methods-and-protocols.html>. Accessed 12 Sep 2022.

Blanco, G., F. Sergio, J. A. Sanchéz-Zapata, J. M. Pérez-García, F. Botella, F. Martínez, I. Zuberogoitia, O. Frías, F. Roviralta, J. E. Martínez, et al. 2012. Safety in numbers? Supplanting data quality with fanciful models in wildlife monitoring and conservation. Biodiversity and Conservation 21:3269–3276.

Boersma, P., and D. Weenink. 2014. Praat: doing phonetics by computer. <http://www.praat.org/>. Accessed 25 Nov 2021.

Brooker, S. A., P. A. Stephens, M. J. Whittingham, and S. G. Willis. 2020. Automated detection and classification of birdsong: an ensemble approach. Ecological Indicators 117:106609.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Burton, A. C., M. K. Sam, C. Balangtaa, and J. S. Brashares. 2012. Hierarchical multi-species modeling of carnivore responses to hunting, habitat and prey in a West African protected area. PLoS ONE 7:e38007.

Buxton, R. T., P. E. Lendrum, K. R. Crooks, and G. Wittemyer. 2018. Pairing camera traps and acoustic recorders to monitor the ecological impact of human disturbance. Global Ecology and Conservation 16:e00493.

Cardoso, L. 2019. Translations and translation gaps: the gunshot acoustic surveillance experiment in Brazil. Sound Studies 5:52–71.

Chacón-Rodríguez, A., P. Julián, L. Castro, P. Alvarado, and N. Hernández. 2011. Evaluation of gunshot detection algorithms. IEEE Transactions on Circuits and Systems I: Regular Papers 58:363–373.

Choi, K. S., M. Librett, and T. J. Collins. 2014. An empirical evaluation: gunshot detection system and its effectiveness on police practices. Police Practice and Research 15:48–61.

Digby, A., M. Towsey, B. D. Bell, and P. D. Teal. 2013. A practical comparison of manual and autonomous methods for acoustic monitoring. Methods in Ecology and Evolution 4:675–683.

Dobbins, M., R. Sollmann, S. Menke, A. Almeyda Zambrano, and E. Broadbent. 2020. An integrated approach to measure hunting intensity and assess its impacts on mammal populations. Journal of Applied Ecology 57:2100–2111.

Freire, I. L., and J. A. Apolinario, Jr. 2010. Gunshot detection in noisy environments. Proceeding of the 7th International Telecommunications Symposium, 6–9 September 2010, Manaus, Brazil.

Fuller, T. K. 1990. Dynamics of a declining white-tailed deer population in north-central Minnesota. Wildlife Monographs 110:3–37.

Garland, L., A. Crosby, R. Hedley, S. Boutin, and E. Bayne. 2020. Acoustic vs. photographic monitoring of gray wolves (*Canis lupus*): a methodological comparison of 2 passive monitoring techniques. Canadian Journal of Zoology 98:219–228.

Gibb, R., E. Browning, P. Glover-Kapfer, and K. E. Jones. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods in Ecology and Evolution 10:169–185.

Harrison, R. D. 2011. Emptying the forest: hunting and the extirpation of wildlife from tropical nature reserves. BioScience 61:919–924.

Hartig, F. 2021. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. <https://cran.r-project.org/package=DHARMa>

Heinicke, S., A. K. Kalan, O. J. J. Wagner, R. Mundry, H. Lukashevich, and H. S. Kühl. 2015. Assessing the performance of a semi-automated acoustic monitoring system for primates. Methods in Ecology and Evolution 6:753–763.

Hill, A. P., P. Prince, E. Piña Covarrubias, C. P. Doncaster, J. L. Snaddon, and A. Rogers. 2018. AudioMoth: evaluation of a smart open acoustic device for monitoring biodiversity and the environment. Methods in Ecology and Evolution 9: 1199–1211.

Holsworth, W. N. 1973. Hunting efficiency and white-tailed deer density. Journal of Wildlife Management 37:336–342.

Hossain, A. N. M., A. Barlow, C. G. Barlow, A. J. Lynam, S. Chakma, and T. Savini. 2016. Assessing the efficacy of camera trapping as a tool for increasing detection rates of wildlife crime in tropical protected areas. Biological Conservation 201:314–319.

Johnson, I., T. Brinkman, K. Britton, J. Kelly, K. Hundertmark, B. Lake, and D. Verbyla. 2016. Quantifying rural hunter access in Alaska. Human Dimensions of Wildlife 21:240–253.

Joly, A., H. Goëau, H. Glotin, C. Spampinato, P. Bonnet, W.-P. Vellinga, J.-C. Lombardo, R. Planqué, S. Palazzo, and H. Müller. 2019. Biodiversity information retrieval through large scale content-based identification: a long-term evaluation. Pages 389–413 *in* N. Ferro, and C. Peters, editors. Information retrieval evaluation in a changing world: lessons learned from 20 years of CLEF. Springer, Cham, Switzerland.

McClintock, B. T., L. L. Bailey, K. H. Pollock, and T. R. Simons. 2010*a*. Experimental investigation of observation error in anuran call surveys. Journal of Wildlife Management 74:1882–1893.

McClintock, B. T., L. L. Bailey, K. H. Pollock, and T. R. Simons. 2010*b*. Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. Ecology 91:2446–2454.

McShea, W. J. 2012. Ecology and management of white-tailed deer in a changing world. Annals of the New York Academy of Sciences 1249:45–56.

Meek, P. D., G. A. Ballard, J. Sparkes, M. Robinson, B. Nesbitt, and P. J. S. Fleming. 2019. Camera trap theft and vandalism: occurrence, cost, prevention and implications for wildlife research and management. Remote Sensing in Ecology and Conservation 5:160–168.

Mennill, D. J., M. Battiston, D. R. Wilson, J. R. Foote, and S. M. Doucet. 2012. Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. Methods in Ecology and Evolution 3:704–712.

Peres, C. A., and E. Palacios. 2007. Basin-wide effects of game harvest on vertebrate population densities in Amazonian forests: implications for animal-mediated seed dispersal. Biotropica 39:304–315.

Piña-Covarrubias, E., A. P. Hill, P. Prince, J. L. Snaddon, A. Rogers, and C. P. Doncaster. 2019. Optimization of sensor deployment for acoustic detection and localization in terrestrial environments. Remote Sensing in Ecology and Conservation 5:180–192.

R Core Team. 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Simons, T. R., M. W. Alldredge, K. H. Pollock, and J. M. Wettroth. 2007. Experimental analysis of the auditory detection process on avian point counts. Auk 124:986–999.

Singh, R. B., H. Zhuang, and J. K. Pawani. 2021. Data collection, modeling, and classification for gunshot and gunshot-like audio events: a case study. Sensors 21:7320.

Stedman, R. C., P. Bhandari, A. E. Luloff, D. R. Diefenbach, and J. C. Finley. 2008. Deer hunting on Pennsylvania's public and private lands: a 2-tiered system of hunters? Human Dimensions of Wildlife 13:222–233.

Stedman, R., D. R. Diefenbach, C. B. Swope, J. C. Finley, A. E. Luloff, H. C. Zinn, G. J. San Julian, and G. A. Wang. 2004. Integrating wildlife and human-dimensions research methods to study hunters. Journal of Wildlife Management 68: 762–773.

Stowell, D., M. D. Wood, H. Pamuła, Y. Stylianou, and H. Glotin. 2019. Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge. Methods in Ecology and Evolution 10:368–380.

Sueur, J., T. Aubin, and C. Simonis. 2008. Seewave: a free modular tool for sound analysis and synthesis. Bioacoustics 18: 213–226.

Sugai, L. S. M., and D. Llusia. 2019. Bioacoustic time capsules: using acoustic monitoring to document biodiversity. Ecological Indicators 99:149–152.

Tumes, K. 2007. Out of my way: using qualitative methods to understand recreation conflict between bushwalkers and mountain bike riders. Anthropological Notebooks 13:45–55.

Van Deelen, T. R., and D. R. Etter. 2003. Effort and the functional response of deer hunters. Human Dimensions of Wildlife 8:97–108.

Wijers, M., A. Loveridge, D. W. Macdonald, and A. Markham. 2021. CARACAL: a versatile passive acoustic monitoring tool for wildlife research and conservation. Bioacoustics 30:41–57.

Wildlife Acoustics. 2016. Song Meter SM4 user guide. Wildlife Acoustics, Inc., Maynard, Massachusetts, USA.

Wrege, P. H., E. D. Rowland, S. Keen, and Y. Shiu. 2017. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. Methods in Ecology and Evolution 8:1292–1301.

Ziegler, S., J. E. Fa, C. Wohlfart, B. Streit, S. Jacob, and M. Wegmann. 2016. Mapping bushmeat hunting pressure in Central Africa. Biotropica 48:405–412.

Zuur, A. F., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer, New York, New York, USA.

Associate Editor: S. Brainerd.

How to cite this article: Hedley, R. W., B. Joubert, H. K. Bains, and E. M. Bayne. 2022. Acoustic detection of gunshots to improve measurement and mapping of hunting activity. Wildlife Society Bulletin 46:e1370. https://doi.org/10.1002/wsb.1370