IBIS
BOU international journal of avian science

Ibis (2024) doi: 10.1111/ibi.13375

Acoustic detection rate can outperform traditional survey approaches in estimating relative densities of breeding waders

DAVID JARRETT* & STEPHEN G. WILLIS
Department of Biosciences, Durham University, South Road, Durham, DH1 3LE, UK

Passive acoustic devices are increasingly being used to monitor biodiversity. However, few studies have compared the accuracy of acoustic surveys and traditional surveys against ground-truthed data. Here, we assess whether acoustic recorders used in conjunction with an artificial intelligence (AI) classifier can predict the relative breeding density of four wader species better than traditional fieldworker transect surveys. In a 27-km^2 upland study site, acoustic data were collected at 83 sampling points and analysed using the BirdNet bird-sound classifier to estimate vocal detection rate at each location; we also carried out concurrent transect bird surveys. To ground-truth these approaches, intensive field surveys were undertaken to identify each breeding territory of our focal species. With both the acoustic dataset and the transect dataset, we used similar analytical approaches (random forest regression trees) to predict relative territory density across the study site, and then compared these predictions with the territory density obtained from the intensive field surveys. The classifier performed well at identifying the presence of target species' vocalizations within 3-s periods for Lapwing (accuracy = 0.911), Curlew (0.826) and Oystercatcher (0.841), but less well for Golden Plover (0.699). For Curlew and Oystercatcher, the predictions obtained from the acoustic approach were a better fit to actual territory density than the transect approach. In contrast, for Lapwing and Golden Plover, the transect predictions outperformed the acoustic predictions, with the acoustic model particularly poor for Golden Plover. We attributed these differences to the performance of the classifier, species' ecology and vocal behaviour. Data gathering for the acoustic approach was more time-efficient than the transect surveys, requiring less than a quarter of the fieldworker days. We conclude that there is high potential for acoustic approaches to augment traditional methods, although species' ecological characteristics should be considered: species that vocalize more frequently, at higher amplitudes and hold larger territories will be better-suited to sampling-based acoustic methods.

Keywords: AI, bioacoustics, breeding waders, long-term monitoring, machine learning.

Passive acoustic devices are increasingly being used to monitor biodiversity (Sugai *et al.* 2019), with methods developed to monitor sounds of cetaceans (Kowarski & Moors-Murphy 2021), fish (Popper & Hawkins 2019), bats (Newson *et al.* 2017), insects (Zilli *et al.* 2014), amphibians (Dutilleux & Curé 2020), canids (Graf & Hatlauf 2021) and birds (Pérez-Granados & Traba 2021). Increased availability of programmable open-source devices based on single-board computers and microcontrollers (Whytock & Christie 2017, Hill *et al.* 2019) allow classification models to run in real-time on devices (Katsis *et al.* 2022). Additionally, developments in machine learning mean that it is feasible to rapidly extract vocalizations of interest from large datasets, drastically reducing the time needed for analysis (Morales *et al.* 2022). For example, the BirdNet sound classifier tool

*Corresponding author.
Email: jarrett.ecology@gmail.com

(Kahl *et al.* 2021) can identify more than 3000 species of European and American birds, and has been used for many ecology and conservation projects (Pérez-Granados 2023).

Acoustic recording units (ARUs) are increasingly being used in long-term monitoring (Lewis *et al.* 2021), although the integration of acoustic methods with existing monitoring approaches may not be straightforward (Doser *et al.* 2021). Many countries have existing long-term monitoring schemes for breeding birds that are carried out by fieldworkers using methods such as transect surveys (Heywood *et al.* 2023) and timed point counts (Sauer *et al.* 2013). Many sources of bias arise from human observers carrying out survey visits, for example the variation in bird behaviour associated with different weather conditions (Hoodless *et al.* 2006), the diel and seasonal timing of survey visits (Thirgood *et al.* 1995), and the skill of the individual observer (Lindenmayer *et al.* 2009). These biases can be mitigated to some extent by the gathering of large datasets and appropriate analytical techniques. Methods that involve standardized ARU deployments are also likely to be subject to biases, particularly related to the performance of ARUs in different habitats and environmental conditions, and the vocal behaviour of target species (Digby *et al.* 2013, Doser *et al.* 2021). However, as the current generation of acoustic recorders can gather very large datasets from one deployment and can be programmed with any temporal sampling regime, some of these biases can be mitigated.

One of the key challenges for the long-term monitoring of biodiversity with ARUs is being able to estimate densities of target species. At present, because most ARUs only have one microphone, assessing directionality or the origin of a sound is very difficult, making it challenging to know how many sound-producing individuals are present at a recording location. Furthermore, using calibrated equipment that is capable of localizing sounds is labour-intensive in terms of analysing data (Rhinehart *et al.* 2020). As such, a range of related approaches to estimating territory density have been developed that use the frequency of calls or detections per recording period at a single recording point as a proxy for territory or species density (Pérez-Granados & Traba 2021, Hutschenreiter *et al.* 2024). The most frequently used is Vocal Activity Rate, which uses the number of calls per time period (Pérez-Granados & Traba 2021); a related approach is detection rate, the rate at which a species vocalization is detected (typically using some kind of automatic classifier that classifies specific time chunks) in specific time periods irrespective of how many individual calls there are (Hutschenreiter *et al.* 2024).

The relationship between a vocal activity index and territory density is likely to vary between sites and species with various factors affecting song rates (Pérez-Granados & Traba 2021). For example, where songbirds breed at higher territory density, higher per-capita song rates and song lengths can occur (Goretskaia 2004). Similarly, unpaired males can sometimes be more vocal than paired males, skewing territory density estimates (Greig-Smith 1982, Amrhein *et al.* 2002). Moreover, the size of a breeding territory and vocalization amplitude will also influence the relationship between vocal activity and territory density, and low breeding productivity may be associated with an earlier cessation of vocal activity in the breeding season (Nebel & McCaffery 2003). For vocal activity indices to be incorporated into long-term monitoring, it is necessary to test not only the relationship between vocal activity derived from ARUs and results from existing survey methods (Vold *et al.* 2017, Pérez-Granados *et al.* 2019), but also to compare both with independent, ground-truthed territory density data. However, few studies include comparisons to ground-truthed data, and where acoustic data are only compared to traditional survey methods, the inferences drawn may be of less value because of the biases in the data obtained from the traditional survey methods.

Whilst acoustic recorders used alongside automatic classifiers are likely to have significant utility across many taxa and habitats, here we explore their potential in open upland landscapes. In such areas, access can be challenging, species diversity and density are typically low, and disturbance can impact the breeding success of some species at critical times of the season. Hence, these landscapes and their species could be well-suited to acoustic monitoring. The UK uplands support important breeding populations of several ground-nesting wading birds, and such species could be particularly suited to acoustic monitoring because they vocalize frequently and at high amplitudes. There were four species of wader sufficiently abundant at the study site to model relative density of territories: Eurasian Oystercatcher

Haematopus ostralegus (hereafter Oystercatcher), Northern Lapwing *Vanellus vanellus* (hereafter Lapwing), European Golden Plover *Pluvialis apricaria* (hereafter Golden Plover) and Eurasian Curlew *Numenius arquata* (hereafter Curlew). There were also Dunlin *Calidris alpina*, Common Snipe *Gallinago gallinago* and Common Sandpiper *Actitis hypoleucos* present in lower densities for which data were not analysed. The four target species (Oystercatcher, Lapwing, Golden Plover and Curlew) have long calls, for example 30 s or longer per call for Oystercatcher or Curlew (see Figs. S1–S4 for example calls).

We hypothesize that using detection rate (positive detections per time period) may be a suitable means of estimating territory density for wader species that vocalize regularly across large areas and have large overlapping territories, such as Curlew and Oystercatcher. We consider it likely that detection rate may outperform transect surveys for these species in terms of the accuracy of the predictive modelling. However, we hypothesize that it will work less well for species – Golden Plover and Lapwing – that vocalize less frequently or across smaller territories. Using detection rate means that some individual data points will be non-independent because each individual call will produce multiple detections; however, because we will generate a very large dataset we do not expect this to unduly bias the data. We also hypothesize that the acoustic approach may require fewer resources than the transect-based approach, as a consequence of the time requirements of transect surveys (particularly in challenging terrain). The objectives of this study are thus to (i) investigate whether detection rate can be used to predict relative territory density for a suite of common upland wading bird species using a predictive modelling approach with environmental covariates (e.g. habitat, elevation); (ii) assess the extent to which detection rate varies in its effectiveness for different wader species; (iii) compare densities derived from detection rate and a traditional transect survey approach to a ground-truthed dataset of breeding wader territories; and (iv) compare the relative resource requirements of the two approaches.

It should be noted that in this paper, we frequently use terms such as ‘territory’ and ‘territory density’. Such terms have many meanings in ecology, yet herein, we use territory to refer to an area where an individual or pair exhibit territorial behaviours during the breeding season associated with a breeding attempt. The study species all have different behaviours during the pre-nesting, nesting and brood-rearing phase: Lapwing nest semi-colonially, Curlew and Oystercatcher have large overlapping territories, and Golden Plover have clearly delineated, non-overlapping territories (Cramp & Simmons 1983).

METHODS

Study site

Delnadamph Estate (Fig. 1) is a traditional upland estate managed primarily for grouse shooting in upper Donside, Aberdeenshire, in the Cairngorms National Park in Scotland. It covers approximately 27 km^2, spanning altitudes from approximately 400 to 830 m asl. On the higher ground (> 700 m) there is a mix of dry lichen heath, blanket bog and some areas of degraded peatland. At intermediate elevations (450–700 m) the primary habitat is heather moorland, managed with rotational burning. On the lower ground, flat expanses of wet grassland and blanket bog fringe the upper reaches of the River Don. There are small areas of plantation woodland and young native woodland,

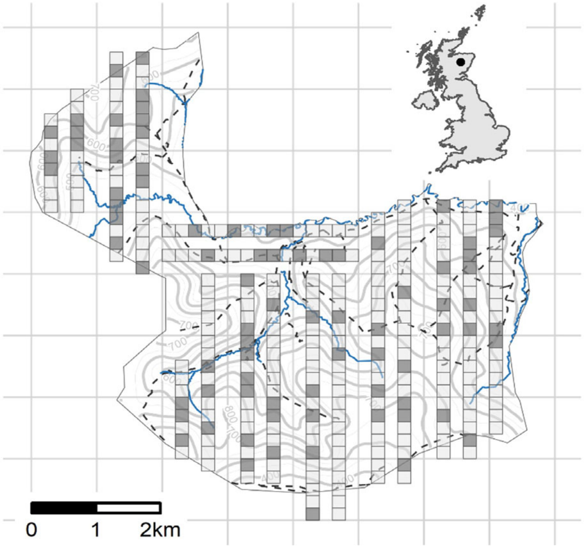

Figure 1. Map of Delnadamph Estate with (inset) the site locality in the UK marked (black circle). Grey contours indicate elevation, blue lines signify rivers and dashed lines indicate access tracks. The 22 surveyed transects are shown divided into the 200-m segments. The segments in grey (n = 83) also had an Audiomoth deployment.

and extensive patches of Juniper *Juniperus communis* with scattered trees alongside smaller streams. The bird communities and ecology of the study site are described in more detail in Jarrett and Bennett (2023).

Data gathering

Transect surveys. Transect surveys for breeding birds, following the survey methods of Heywood *et al.* (2024), were conducted across the study site (Fig. 1) with all bird species encountered recorded. Surveys began shortly after dawn and were completed by approximately 10:30 AM. They were carried out in dry, relatively calm and clear weather conditions. Most transects were 3–4 km in length. Two visits were carried out to each transect apart from one that only had one visit owing to poor weather. Early visits were carried out between 21 April and 13 May, and late visits between 7 June and 22 June.

All bird species were recorded, although here we only consider data on the most abundant wader species: Oystercatcher, Lapwing, Golden Plover and Curlew. All individuals were recorded whether seen or heard, and activity or sex was not recorded. Birds were allocated to a distance band (0–25 m; 25–50 m; 50–100 m) perpendicular to the transect line (records further than 100 m either side of the transect line were excluded) and to a 200-m segment of the transect based on the location of the first observation. There were 287 segments surveyed in total. No attempt was made to separate records into breeding territories while carrying out transect surveys.

Breeding wader territory mapping. To produce territory maps with which to ground-truth the breeding wader data, all areas with breeding waders present were visited at least three times across the breeding season, and the estimated centre of each wader territory was recorded by watching from vantage points or walking through suitable habitat and recording territorial birds or nest locations and estimating the centre of the breeding territory from records across all visits. The territory mapping was carried out independently of the transect surveys. We consider this dataset to be close to the actual number of territories across the study site, although for all these species over- or under-counting is possible. Across all four species, only four additional territories were discovered on third visits (or later), with third and later visits mostly spent determining the (approximate) centre of territories.

Acoustic recorders. Thirty-four Audiomoth Acoustic recorders (v.1.1.0 and v.1.2.0, Open Acoustic Devices, Oxford, UK) were used in the study site. With data gathered from 83 different randomly selected locations, each recorder location was on a randomly selected perimeter of a randomly selected 200-m × 200-m segment which was also covered by the transect surveys. The acoustic recorders were placed to face the centre of the segment. All audio recorders were deployed in a location for a minimum of 10 days, but some were *in situ* for longer as a result of the practicalities of retrieving them from remote locations. Following data retrieval, recorders were re-deployed in new locations. The earliest recording date was 17 April, and the last recording date was 12 June. The audio recorders were deployed fixed to small wooden stakes approximately 30 cm above ground, or higher if the vegetation was higher than 30 cm.

Recorders were programmed to record for 5 min in every 30 min between the times of 3:00 AM and 9:00 AM on each day that they were active (initial testing had found that this was approximately the period of highest vocal activity for the target species). Recordings were made at 16 kHz frequency. The recorders were housed in standard Audiomoth IPX7 waterproof cases (Open Acoustic Devices).

Data analysis

Explanatory variables

The study area was divided into 825 segments measuring 200 m × 200 m (subsets of these segments were covered by the transect surveys and acoustic recorders as described above). For habitat cover (Fig. 2), satellite images (Google Earth, accessed 05/07/2023) were used in QGIS (QGIS.org 2024) to classify habitats within the study area as either blanket bog, rotationally burnt heather moorland, unburnt heath, degraded peat, grassland, juniper scrub, woodland or open water, by drawing polygons around areas of each habitat type. For each 200-m × 200-m segment, the extent of each habitat type was allocated to one of four classes (< 1% = class 0; 1–10% = class 1;

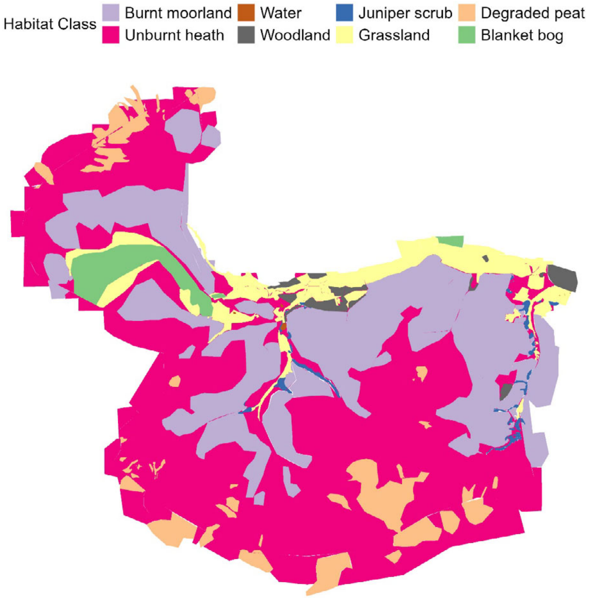

Figure 2. Habitat map of the 27-km^2 study site. The eight different habitat classes identified from satellite images are coloured as described in the key.

10–50% = class 2; > 50% = class 3). For each segment, mean elevation (Fig. 3a) and mean slope (Fig. 3b) were extracted from the NASADEM digital elevation model. Easting and Northing were also extracted for each segment for use as explanatory variables in models to account for spatial autocorrelation.

Acoustic data

Recordings were first analysed using the BirdNet analyser tool v.2.2 (Kahl *et al.* 2021). BirdNet divides recordings into 3-s clips and classifies the recordings using a Deep Artificial Neural Network, previously trained on a large dataset of acoustic samples, with overlap set to 0.0, sensitivity to 1.0 and minimum confidence to 0.1 (Birdnet default values (Kahl *et al.* 2021); initial tests found that altering these values did not significantly improve the dataset). The species list was initially set to include all species for which breeding was known or considered possible, though the only four species for which data were analysed were Golden Plover, Lapwing, Oystercatcher and Curlew.

The accuracy of the BirdNet classifier was validated by manually labelling a subset of 200 randomly selected 5-min recordings. We viewed the acoustic data in Raven Pro v.1.6.5 (Cornell Lab of Ornithology 2024), using a 1722-point Hamming window spectrogram for visualization. All vocalizations by wader species were labelled by drawing a bounding box around the smallest area possible surrounding each vocalization. Only vocalizations that were audible and produced a clear spectrogram were labelled. In cases where vocalizations of either the same species or different target species overlapped, a separate bounding box was drawn for each separable vocalization. One bounding box was drawn for bursts of calls with consistent intermediate periods. For complex, extended vocalizations, such as a bubbling Curlew display call, the bounding box was drawn around the whole of the vocalization.

The approach was assessed using two complementary approaches. First, for each 3-s period a confusion matrix was used to assess the precision

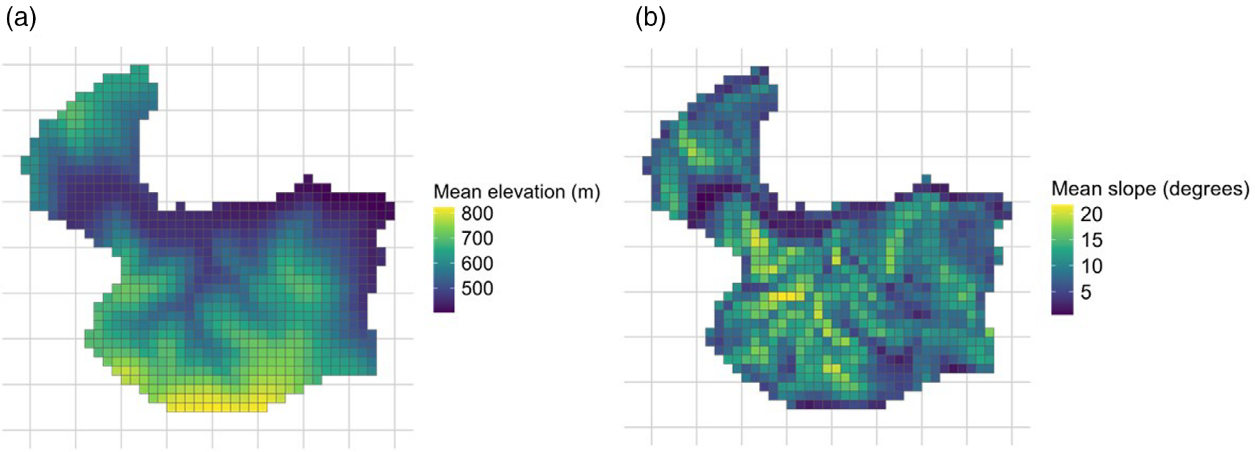

Figure 3. (a) Mean elevation and (b) mean slope in each 200-m × 200-m segment of the study site with a 1-km × 1-km grid overlay.

and accuracy of the BirdNet classifier against the manually labelled dataset. Second, for each species a linear regression was used to assess the accuracy of BirdNet at estimating vocalization frequency for the 200 manually labelled 5-min recordings. The response variable was the total time (within the 5-min recording) of positive manual labels, and the explanatory variable was the total time of positive BirdNet labels (applying the 0.1 confidence threshold). R-squared (R^2) values were used to compare the accuracy of BirdNet-predicted vocalization frequency for 5-min periods for each species.

Territory density maps

To assess the performance of the acoustic and transect models against empirical data, we produced territory density maps for each species based on the independent territory mapping exercise. Breeding waders typically vocalize and perform display flights across a larger area than the immediate nest-site, but with variation in territory size in different habitats (Ewing *et al.* 2017, Bowgen *et al.* 2022). Because limited information exists on territory size for each species, territory density maps were produced using quartic kernel densities with radii 250 m, 500 m, 750 m, 1000 m and 1250 m around the centre-point of each territory (see below).

Predictive modelling

For both the acoustic data and the transect data we used random forest regression trees to make predictions of territory density across the study area from the sampled data. The predictive outputs of these two modelling approaches were then compared to the wader territory maps. Random forests were used for the predictive modelling because they have good predictive power (Evans *et al.* 2011, Mi *et al.* 2017), are robust to overfitting and rely on few assumptions about the distribution of explanatory variables, and collinearity of explanatory variables is less problematic than with linear modelling approaches (Howard *et al.* 2014). Data analysis was carried out in R v.3.6.1 (R Core Team 2022) using the packages *randomForest*, *rgdal*, *sf* and *dplyr*.

For the transect data, random forest regression trees (with default settings) were used to predict the relative abundance of each of the four target species of breeding wader across the study site in the 200-m × 200-m segments. We used the count of individuals up to 100 m from the transect line on each segment as the response variable (we ignored distance bands as there were too few records for some species to produce distance functions) and the environmental covariates (see above) as explanatory variables (including Easting and Northing).

With the acoustic dataset, for each of the 83 deployment sites 100 × 5-min samples were randomly selected from between 3:00 AM and 9:00 AM during the period of deployment, excluding those samples significantly affected by wind or rain. The number of samples analysed was restricted to 100 because there was wide variation in the number of samples obtained from each recorder owing both to the challenge of recovering recorders from remote locations and variable battery life (rechargeable batteries were used which hold less charge than non-rechargeable lithium batteries). These samples were analysed with the Birdnet classifier to identify all segments with a positive classification for a target species to produce a binomial variable (presence or absence of the target species in each 3-s segment), which was the response variable in the random forest model. Each recorder location was assigned to the 200-m × 200-m segment in which it was positioned, and the environmental explanatory variables for each segment were then used in the model. Time and date were also included as explanatory variables, time being the mid-point of the 5-min period and date as the number of days since 1 April. The models then predicted the likelihood of positive classifications in each 200-m × 200-m segment of the study site.

For both the acoustic and transect approaches, the number of random forest regression trees (RFRTs) in each random forest was 500, and the number of variables sampled as candidates for each tree was three. In both cases, all explanatory variables were retained in all models, as the main purpose of the modelling was to compare the predictive power of the acoustic data versus the transect data, rather than to assess the effect of environmental variables on wader density. Model performance was assessed using 'out-of-the-bag' predictions, meaning that estimates of predictive power are based on independent data. For both models, the proportion of variation in the response variable (positive 3-s classifications in the acoustic data or count data from transects) explained by the environmental variables was assessed using R^2.

The predictive power of the acoustic and transect models were assessed against the independently assessed territory density maps. We ran linear models with actual territory density in each segment as the response variable, and the random forest predictions from the acoustic and transect approaches as the explanatory variable. A 2nd order polynomial term was included in the models to account for the possibility of non-linearity. For each species, separate models were run for the acoustic and transect models and for each of the five territory size categories to identify: (i) whether the acoustic or transect random forest was the best fit to territory density, and (ii) which assumed territory size was the best fit for each species.

RESULTS

The independent territory mapping identified 28 territories for Oystercatcher, as well as Lapwing (51), Golden Plover (27) and Curlew (47) territories in the study site. On the fieldworker transect surveys, there were 27 encounters with Oystercatcher, 59 with Lapwing, 57 with Golden Plover and 73 with Curlew. The 100 × 5-min samples from each of the 83 acoustic recorder deployments generated 830 000 × 3-s segments for classification using the BirdNet classification tool. The classifier identified 17 421 × 3-s segments with Curlew vocalizations, 6916 with Oystercatcher, 6205 with Lapwing and 531 with Golden Plover. An approximate estimate of vocalization frequency per territory across the site was produced by dividing the number of classifications by the total number of territories identified in the independent territory mapping. This resulted in 1112 s of vocalization per territory for Curlew, 741 s per Oystercatcher territory, 365 s per Lapwing territory and 59 s per Golden Plover territory.

Using the manually labelled 3-s segments as a validation dataset, BirdNet Analyser was better at classifying Oystercatcher (accuracy = 0.84, precision = 0.84, recall 0.19), Lapwing (accuracy = 0.91, precision = 0.86, recall = 0.17) and Curlew (accuracy = 0.83, precision = 0.93, recall = 0.18) than for Golden Plover (accuracy = 0.70, precision = 0.55, recall = 0.01). Across the 200 manually labelled 5-min periods (Fig. 4), BirdNet Analyser produced values of summed vocalization that were highly correlated to the manually labelled datasets for three of the four species. For Oystercatcher, the coefficient of the linear model (where 1 would be perfect alignment and the intercept is fixed at 0) was 0.840 and $R^2 = 0.900$, for Curlew the coefficient was 0.782 and $R^2 = 0.899$, for Lapwing the coefficient was 1.083 and $R^2 = 0.893$, but for Golden Plover the coefficient was 0.036 and $R^2 = 0.238$.

For Oystercatcher ($R^2 = 0.594$) and Curlew ($R^2 = 0.599$) the acoustic models performed well in explaining variance of detection rate between recorder locations, more so than the equivalent test of the transect models ($R^2 = 0.448$ and 0.301, respectively). For Golden Plover and Lapwing, both the acoustic models ($R^2 = 0.165$ and 0.205, respectively) and transect models ($R^2 = 0.270$ and 0.196, respectively) performed relatively poorly.

For each species, we used the independent ground-truthed territory maps to produce kernel density maps for five assumed territory sizes (250–1250 m). We used linear models, with a polynomial term to account for non-linearity, to assess how accurately the acoustic and transect random forest models predicted actual territory density (for the five different assumed territory sizes (Table 1)). The best-fitting territory size was larger for the acoustic models than for the transect models for Curlew (acoustic: 1000 m, transect: 750 m), Oystercatcher (acoustic: 1250 m, transect: 500 m) and Golden Plover (acoustic: 1000 m, transect: 500 m), while for Lapwing there was no difference (acoustic: 750 m, transect 750 m). The best acoustic model outperformed the best transect model for Oystercatcher and Curlew, while for Golden Plover and Lapwing, the transect models performed better (Table 1). The relative density predictions obtained from the acoustic modelling are a good approximation of the distribution of territories across the study site (Figs. S1–S4). We found evidence of a saturation effect (Oppel *et al.* 2014), where, at higher territory densities, increases in territory density did not result in equivalent increases in the frequency of detected vocalizations (Figs. S5–S8).

In terms of resource requirements, to program, deploy and collect recorders from the 83 locations (in a study site relatively well covered by off-road tracks, see Fig. 1) was approximately four fieldwork days. Had there been no need to move the recorders between sites (e.g. if 83 recorders had been used) then it is likely that the deployment and collection could have been completed in half this time. Carrying out the transect surveys took 18 fieldwork days with good weather. Creating the

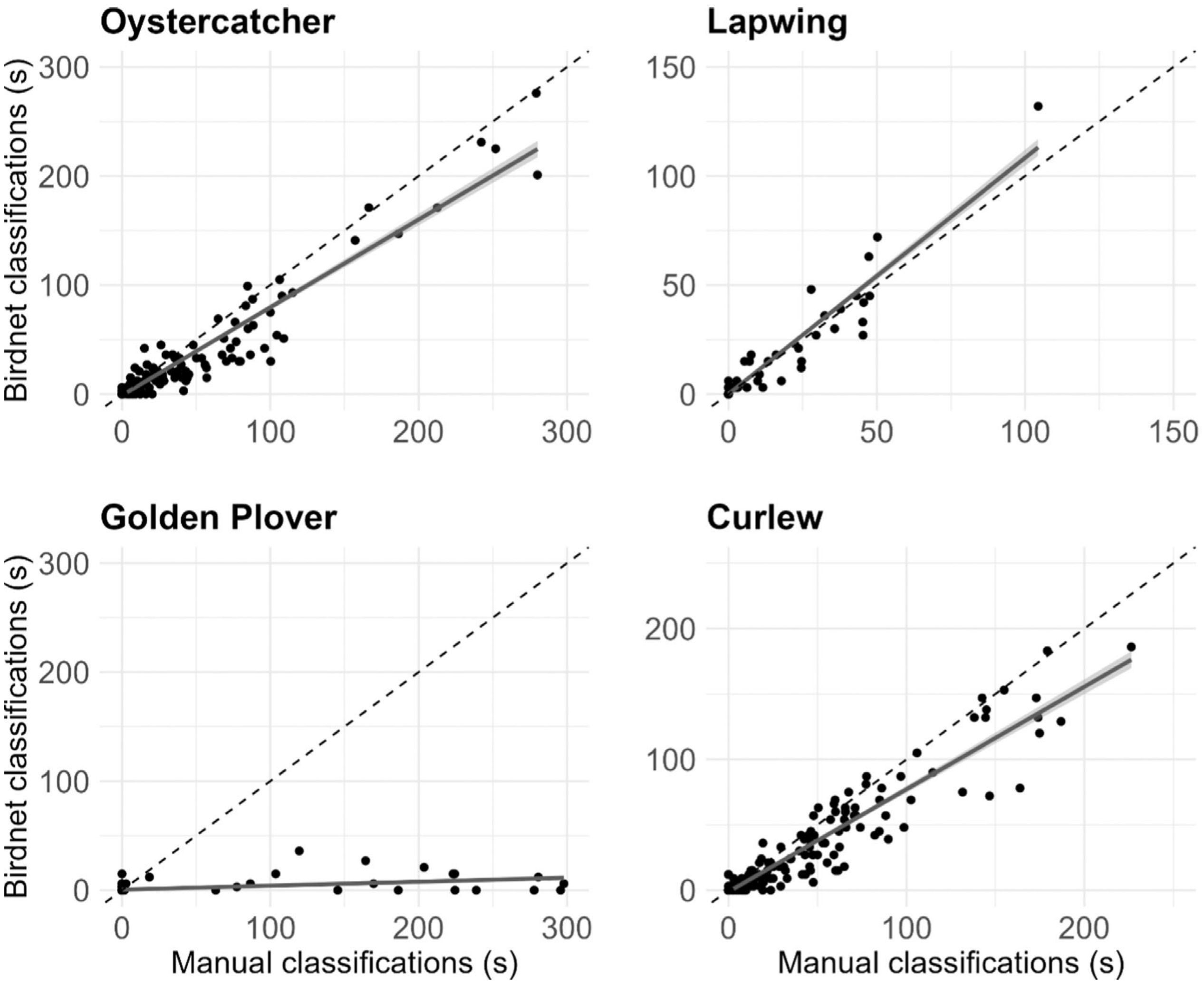

Figure 4. Linear models assessing the accuracy of BirdNet at estimating vocal activity in a 5-min period. For 200 manually labelled 5-min periods, the total amount of labels (in s) for each species are plotted against BirdNet classifications (in s). A linear regression (grey line) with 95% confidence intervals (light grey shading) is plotted.

Table 1. Comparison of the performance of acoustic and transect models in predicting territory density across the study site with different assumed territory sizes applied to the ground-truthed territory maps for each species.

Model	Acoustic					Transect				
Territory size (radius)	250 m	500 m	750 m	1000 m	1250 m	250 m	500 m	750 m	1000 m	1250 m
Oystercatcher	0.420	0.502	0.527	0.535	***0.537***	0.434	0.471	0.492	0.478	0.459
Golden Plover	0.0954	0.179	0.225	0.242	0.231	0.285	***0.414***	0.412	0.387	0.311
Lapwing	0.243	0.379	0.429	0.422	0.356	0.313	0.444	***0.466***	0.458	0.366
Curlew	0.373	0.595	0.669	***0.684***	0.676	0.340	0.538	0.594	0.587	0.558

R^2 values are shown for acoustic and transect models for territory density grids calculated using 250 m, 500 m, 750 m, 1000 m and 1250 m kernels. The R^2 for the best-fitting model for each species across all of the territory sizes and the acoustic and transect models is shown in bold and italics.

manually labelled dataset and then running analyses to validate the acoustic approach took approximately 10 days. Subsequent analyses took similar amounts of time for each approach.

DISCUSSION

This study demonstrates that, for some species of breeding waders, using acoustic recorders and an automatic classifier can predict relative density of breeding territories more accurately than a traditional transect survey-based approach. Consistent with our hypotheses, for Oystercatcher and Curlew the acoustic-based approach clearly outperformed the transect-based approach, whereas for Golden Plover and Lapwing the transect-based approach performed better, although for these two species both approaches were poor at predicting relative territory density. Additionally, our hypothesis that the resource requirement to gather and analyse the acoustic data would be less than for the transect surveys was confirmed. These findings demonstrate the potential for acoustic methods to augment traditional monitoring approaches, but also show that closely related species can vary in the extent to which they are suited to acoustic methods.

Predictive modelling approach

We used a predictive modelling approach to compare the effectiveness of acoustic recorders with transect surveys at predicting the relative density of breeding territories across the study site. We used this approach because the breeding territories of our study species are much larger than the 200-m segments (Bowgen *et al.* 2022) and the vocal activity or movement of the target species associated with a breeding territory will not be limited to the 200-m segments, and territorial vocalization behaviour can overlap across multiple territories. This meant that an assumption that the acoustic data would correlate to the presence or absence of territories at the 200-m × 200-m segment size would not be justified, and so we used a range of territory buffers around each estimated territory centre to create a territory density map. Moreover, there is strong evidence that the distribution of breeding waders in open upland habitats is highly associated with slope, habitat and elevation (e.g. see Stillman & Brown 1994) – environmental variables for which we had good data, so we assumed that a predictive modelling approach would be an effective way to compare the acoustic and transect datasets.

Ecological characteristics of target species

The vocal behaviour of the target species influenced the effectiveness of the acoustic approach – the total number of Curlew positive classifications per territory was approximately 19 times higher than for Golden Plover – meaning that the Curlew acoustic model predictions were based on far more positive data points than the Golden Plover acoustic model. Golden Plover nests were at higher elevations and in more exposed areas than Curlew so some of this difference may be driven by the difficulty in getting clear recordings in higher altitude, exposed areas, and consequently the classifier accuracy was low for Golden Plover. Golden Plover are specifically known to have low detectability during the incubation period, but peaks of detectability occur during the territory-establishment and brood-rearing periods (Byrkjedal & Thompson 1998, Pearce-Higgins & Yalden 2005). On the transect surveys, Curlew were recorded only slightly more frequently than Golden Plover (73 times vs 57 times, respectively). For Golden Plover, vocal activity is very limited around the nest-site after the incubation period starts, except in response to intruders/predators, and there were instances of non-incubating birds calling frequently on lower ground (some distance from breeding territories), confounding the assumption of the acoustic modelling approach that more vocal activity is recorded around breeding territories. In contrast to Golden Plover, the other three wader species vocalized throughout the breeding season, breed in less exposed areas and are less likely to vocalize away from the breeding territory. As a result, the acoustic recorder and automatic classifier approach was more effective. For Lapwing, the lower frequency of vocalization per territory and the fact that they nest in close proximity – often in high-density clusters – may limit the effectiveness of both the acoustic and transect models, although the greater spatial coverage of the transect surveys may be better suited to this species.

We did not have *a priori* knowledge of how large the acoustic footprint of each breeding territory would be for each target species: it is likely to be influenced by the amplitude of vocalizations, how far from territory centres birds vocalize, the topography of the study site, weather conditions and also the effective detection range of the sensors (Thomas *et al.* 2020). To account for these effects, we produced ground-truthing maps with five different assumed territory sizes to see which would fit the predictive acoustic models best. The acoustic models were a better fit for larger assumed territory sizes for three of the four species: acoustic sensors may effectively survey a larger area than fieldworker transects and so register detections from birds further away from territory centres. Curlew and Oystercatcher were a better fit for larger territory sizes – presumably because these species make large display flights vocalizing far beyond the immediate nest-site, meaning the acoustic footprint of a territory is large. In contrast, Lapwing and Golden Plover had smaller acoustic territories as vocalizations do not carry as far. Lapwing also nest in very close proximity to one another in colonies of (at this study site) up to eight pairs in one field (approximately 2 ha). Species distribution models may work much less well for colonial species (Engler *et al.* 2017) because effective sample size is reduced (Wisz *et al.* 2008). We did not adjust the model to account for Lapwing semi-coloniality to ensure that the approach used for each species was the same, and because our aim was to compare the extent to which modelled acoustic data predict territory density given different species traits.

As territory density increased for each species, vocal activity did not increase equivalently (see Figs. S1–S4) and reasons for this may be two-fold: there may be genuine decreases in vocal activity per pair at higher densities and it may also be the case that birds become more likely to vocalize simultaneously (Oppel *et al.* 2014). Particularly for Curlew, the recorders in the highest density areas often had two or three birds vocalizing simultaneously.

Classifier performance

Evaluated at the 5-min period scale (Fig. 4), the automatic classifier performed well compared to manually labelled data. For all species precision was relatively high (few false positives) but recall was relatively low (meaning a high number of false negatives) evaluated at the 3-s period scale. This is likely to be partly because, for the manual labels, bounding boxes were drawn precisely around the area of vocalization on the spectrogram, so there were occasions where a small portion of a label overlapped into a 3-s BirdNet classification period, resulting in a positive record in the manual dataset and a negative record in the BirdNet dataset (because the classifier was unable to make a positive ID from a small portion of a vocalization). Indeed, while the random forest model predicts relative territory density best for Curlew, the balanced accuracy of the BirdNet classifier for Curlew is lower than for Lapwing and Oystercatcher; this reiterates that the suitability of a species to acoustic monitoring is not solely a function of the performance of an automatic classifier.

Spatial and temporal sampling

In the acoustic dataset, for each recorder location, 100 × 5-min samples were randomly selected from between 3:00 AM and 9:00 AM, excluding those samples significantly affected by wind or rain. To maximize use of recorders, recorders were moved to different locations and so data were gathered from different locations at different times during the breeding period. While vocal activity will vary across the wader breeding period (Nebel & McCaffery 2003), we considered that sampling different locations at different times during the season would not unduly influence the dataset, providing that there was no spatial or altitudinal bias in the locations surveyed earlier or later in the breeding season. Having more recorders and not needing to move them would have improved coverage and reduced deployment/retrieval time. This may be an additional factor in the poor performance of the Golden Plover acoustic model, because Golden Plover are typically less vocally active than the other species, and vocal activity can vary substantially with stage of breeding (Pearce-Higgins & Yalden 2005). We limited each recorder location to 100 × 5-min samples because for some recorders the batteries failed quite soon after deployment (within 7 days) but we were still able to extract 100 samples from these recorders. For all four target species, the estimated detection rate at individual sampling points would not have been significantly different with more samples. In contrast to the acoustic dataset, in the transect dataset there are effectively only two temporal sampling points (from each of the two survey visits). Because the transect surveys started soon after dawn and ended at around 10:00 AM, different areas were surveyed at times with significantly different levels of bird activity, although this effect is partly mitigated by reversing the order in which the transects are surveyed on first and second visits. In terms of spatial sampling, the transect approach had approximately 3.5-fold greater coverage with 287 segments covered by transects compared to 83 acoustic sampling segments.

Relative resource requirements

When selecting to use an acoustic or transect-based approach, practical and financial concerns will be important. The cost of an individual Audiomoth is currently US$97 (October 2023). The recorders used in this project had been used in two field seasons already: four of the 35 original recorders no longer functioned owing to various problems, and two more failed during the current project breeding season. This level of attrition, in a project with a relatively large number of recorders, is not likely to have had a material effect on the final dataset. The temporal requirements of the acoustic approach were 14 days while the transect approach required 18 days. The subsequent time requirements in terms of analysis for the acoustic and transect data were similar, meaning that, in total, the acoustic approach was less time-consuming than the transect approach, although there were also additional data storage costs associated with the acoustic approach. If this site were to be re-surveyed in future years using an acoustic approach with the same automatic classifier and the same target species (and confusion species) then redoing the manually labelled dataset would be unnecessary, so the time requirements would be very low. There are also possibilities for using citizen science to create validation datasets for bioacoustics projects (Jäckel *et al.* 2021).

Future directions and implications

There is significant potential for passive acoustic methods to augment traditional survey methods. However, this study demonstrates that closely related species with relatively similar ecologies can vary significantly in the extent to which they are suited to acoustic monitoring. Researchers should thus proceed with a degree of caution, and must consider factors such as breeding ecology and the vocal behaviour of a species when considering an acoustic approach. An important next step, particularly for breeding waders, will be to understand the extent to which site-specific variables significantly influence vocalization rates – it is likely that factors such as the frequency of anthropogenic disturbance (Diepstraten & Willie 2021) or noise (Cretois *et al.* 2024), predator densities (Lourenço *et al.* 2013), habitat type or food availability will influence vocal activity. Acoustic methods may also be capable of providing information not just on breeding density, but also on wader breeding productivity, which is very important for conservation decision-making (Jarrett *et al.* 2024). More broadly, using acoustic monitoring to estimate absolute territory density across different sites, contexts and management regimes will only be possible with a robust understanding of the relative importance of site-specific factors. More studies such as this, which ground-truth approaches to empirical data, will be valuable in assessing the wider efficacy of acoustic monitoring approaches.

Staff at Delnadamph Estate led by Michael Bennett provided access and support in carrying out surveys and deploying recording devices. Alison Johnston and Aleksi Lehikoinen reviewed a draft of the manuscript and kindly provided many helpful comments. Ruedi Nager, Juan Navedo and two anonymous reviewers at *Ibis* significantly improved the quality of the manuscript with insightful comments and suggestions.

FUNDING

D. J. was supported by a Leverhulme Trust Doctoral Studentship Grant DS-2017-070 (Interdisciplinary Understanding for a Changing Arctic) under the supervision of S. G. W.

AUTHOR CONTRIBUTIONS

David Jarrett: Conceptualization; writing – original draft; writing – review and editing; methodology; formal analysis. **Stephen G. Willis:** Writing – review and editing; supervision.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

ETHICAL NOTE

None.

DATA AVAILABILITY STATEMENT

Data are available upon request to the author at jarrett.ecology@gmail.com.

REFERENCES

Amrhein, V., Korner, P. & Naguib, M. 2002. Nocturnal and diurnal singing activity in the nightingale: correlations with mating status and breeding cycle. *Anim. Behav.* **64**: 939–944.

Bowgen, K.M., Dodd, S.G., Lindley, P., Burton, N.H.K. & Taylor, R.C. 2022. Curves for curlew: identifying curlew breeding status from GPS tracking data. *Ecol. Evol.* **12**: e9509.

Byrkjedal, I. & Thompson, D. 1998. *Tundra Plovers: the Eurasian, Pacific and American Golden Plovers and Grey Plover*, 1st edn. London: T & AD Poyser.

Cornell Lab of Ornithology 2024. *Raven Pro: Interactive Sound Analysis Software (Version 1.6.5)*. Ithaca, NY: The Cornell Lab of Ornithology. Available at: https://www.ravensoundsoftware.com/

Cramp, S. & Simmons, K.E.L. 1983. *The Birds of the Western Palearctic. Vol. III. Waders to Gulls*. Oxford: Oxford University Press.

Cretois, B., Bick, I.A., Balantic, C., Gelderblom, F.B., Pávon-Jordán, D., Wiel, J., Sethi, S.S., Betchkal, D.H., Banet, B., Rosten, C.M. & Reinen, T.A. 2024. Snowmobile noise alters bird vocalization patterns during winter and pre-breeding season. *J. Appl. Ecol.* **61**: 340–350.

Diepstraten, J. & Willie, J. 2021. Assessing the structure and drivers of biological sounds along a disturbance gradient. *Glob. Ecol. Conserv.* **31**: e01819.

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

Doser, J.W., Finley, A.O., Weed, A.S. & Zipkin, E.F. 2021. Integrating automated acoustic vocalization data and point count surveys for estimation of bird abundance. *Methods Ecol. Evol.* **12**: 1040–1049.

Dutilleux, G. & Curé, C. 2020. Automated acoustic monitoring of endangered common spadefoot toad populations reveals patterns of vocal activity. *Freshw. Biol.* **65**: 20–36.

Engler, J.O., Stiels, D., Schidelko, K., Strubbe, D., Quillfeldt, P. & Brambilla, M. 2017. Avian SDMs: current state, challenges, and opportunities. *J. Avian Biol.* **48**: 1483–1504.

Evans, J.S., Murphy, M.A., Holden, Z.A. & Cushman, S.A. 2011. Modeling species distribution and change using random Forest. In Drew, C.A., Wiersma, Y.F. & Huettmann, F. (eds) *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*: 139–159. New York, NY: Springer.

Ewing, S.R., Scragg, E.S., Butcher, N. & Douglas, D.J.T. 2017. GPS tracking reveals temporal patterns in breeding season habitat use and activity of a globally near threatened wader, the Eurasian curlew. *Wader Study* **124**: 206–214.

Goretskaia, M.I. 2004. Song structure and singing behaviour of willow warbler *Phylloscopus trochilus acredula* in populations of low and high density. *Bioacoustics* **14**: 183–195.

Graf, L. & Hatlauf, J. 2021. Distance estimation of howling golden jackals (*Canis aureus*) using relative sound level. *Mamm. Res.* **66**: 567–572.

Greig-Smith, P.W. 1982. Song-rates and parental care by individual male stonechats (*Saxicola torquata*). *Anim. Behav.* **30**: 245–252.

Heywood, J.J.N., Massimino, D., Balmer, D.E., Kelly, L., Marion, S., Noble, D.G., Pearce-Higgins, J.W., White, D.M., Woodcock, P., Wotton, S. & Gillings, S. 2024. The breeding bird survey 2023. In *BTO Research Report 765*. Thetford: British Trust for Ornithology.

Hill, A.P., Prince, P., Snaddon, J.L., Doncaster, C.P. & Rogers, A. 2019. AudioMoth: a low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX* **6**: e00073.

Hoodless, A.N., Inglis, J.G. & Baines, D. 2006. Effects of weather and timing on counts of breeding snipe *Gallinago gallinago*. *Bird Study* **53**: 205–212.

Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D. & Willis, S.G. 2014. Improving species distribution models: the value of data on abundance. *Methods Ecol. Evol.* **5**: 506–513.

Hutschenreiter, A., Andresen, E., Briseño-Jaramillo, M., Torres-Araneda, A., Pinel-Ramos, E., Baier, J. & Aureli, F. 2024. How to count bird calls? Vocal activity indices may provide different insights into bird abundance and behaviour depending on species traits. *Methods Ecol. Evol.* **15**: 1071–1083.

Jäckel, D., Mortega, K.G., Sturm, U., Brockmeyer, U., Khorramshahi, O. & Voigt-Heucke, S.L. 2021. Opportunities and limitations: a comparative analysis of citizen science and expert recordings for bioacoustic research. *PLoS One* **16**: e0253763.

Jarrett, D. & Bennett, M. 2023. Breeding bird and mountain hare survey data from a Scottish drive grouse moor. *Scottish Birds* **43**: 291–304.

Jarrett, D., Lehikoinen, A. & Willis, S. 2024. Monitoring wader breeding productivity. *Ibis* **166**: 780–800.

Kahl, S., Wood, C.M., Eibl, M. & Klinck, H. 2021. BirdNET: a deep learning solution for avian diversity monitoring. *Eco. Inform.* **61**: 101236.

Katsis, L.K.D., Hill, A.P., Piña-Covarrubias, E., Prince, P., Rogers, A., Patrick Doncaster, C. & Snaddon, J.L. 2022. Automated detection of gunshots in tropical forests using convolutional neural networks. *Ecol. Indic.* **141**: 109128.

Kowarski, K.A. & Moors-Murphy, H. 2021. A review of big data analysis methods for baleen whale passive acoustic monitoring. *Mar. Mamm. Sci.* **37**: 652–673.

Lewis, R.N., Williams, L.J. & Gilman, R.T. 2021. The uses and implications of avian vocalizations for conservation planning. *Conserv. Biol.* **35**: 50–63.

Lindenmayer, D.B., Wood, J.T. & MacGregor, C. 2009. Do observer differences in bird detection affect inferences from large-scale ecological studies? *Emu* **109**: 100–106.

Lourenço, R., Goytre, F., del Mar Delgado, M., Thornton, M., Rabaça, J.E. & Penteriani, V. 2013. Tawny owl vocal activity is constrained by predation risk. *J. Avian Biol.* **44**: 461–468.

Mi, C., Huettmann, F., Guo, Y., Han, X. & Wen, L. 2017. Why choose random Forest to predict rare species distribution with few samples in large undersampled areas?

Three Asian crane species models provide supporting evidence. *PeerJ* **5**: e2849.

Morales, G., Vargas, V., Espejo, D., Poblete, V., Tomasevic, J.A., Otondo, F. & Navedo, J.G. 2022. Method for passive acoustic monitoring of bird communities using UMAP and a deep neural network. *Eco. Inform.* **72**: 101909.

Nebel, S. & McCaffery, B.J. 2003. Vocalization activity of breeding shorebirds: documentation of its seasonal decline and applications for breeding bird surveys. *Can. J. Zool.* **81**: 1702–1708.

Newson, S.E., Evans, H.E., Gillings, S., Jarrett, D., Raynor, R. & Wilson, M.W. 2017. Large-scale citizen science improves assessment of risk posed by wind farms to bats in southern Scotland. *Biol. Conserv.* **215**: 61–71.

Oppel, S., Hervias, S., Oliveira, N., Pipa, T., Silva, C., Geraldes, P., Goh, M., Immler, E. & McKown, M. 2014. Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. *Nat. Conserv.* **7**: 1–13.

Pearce-Higgins, J.W. & Yalden, D.W. 2005. Difficulties of counting breeding Golden plovers *Pluvialis apricaria*. *Bird Study* **52**: 339–342.

Pérez-Granados, C. 2023. BirdNET: applications, performance, pitfalls and future opportunities. *Ibis* **165**: 1068–1075.

Pérez-Granados, C. & Traba, J. 2021. Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. *Ibis* **163**: 765–783.

Pérez-Granados, C., Bota, G., Giralt, D., Barrero, A., Gómez-Catasús, J., Bustillo-De La Rosa, D. & Traba, J. 2019. Vocal activity rate index: a useful method to infer terrestrial bird abundance with acoustic monitoring. *Ibis* **161**: 901–907.

Popper, A.N. & Hawkins, A.D. 2019. An overview of fish bioacoustics and the impacts of anthropogenic sounds on fishes. *J. Fish Biol.* **94**: 692–713.

QGIS.org 2024. *QGIS Geographic Information System*. QGIS Association.

R Core Team 2022. *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rhinehart, T.A., Chronister, L.M., Devlin, T. & Kitzes, J. 2020. Acoustic localization of terrestrial wildlife: current practices and future opportunities. *Ecol. Evol.* **10**: 6794–6818.

Sauer, J.R., Link, W.A., Fallon, J.E., Pardieck, K.L. & Ziolkowski, D.J. 2013. The North American breeding bird survey 1966–2011: summary analysis and species accounts. *N. Am. Fauna* **79**: 1–32.

Stillman, R.A. & Brown, A.F. 1994. Population sizes and habitat associations of upland breeding birds in the south Pennines, England. *Biol. Conserv.* **69**: 307–314.

Sugai, L.S.M., Silva, T.S.F., Ribeiro, J.W. Jr. & Llusia, D. 2019. Terrestrial passive acoustic monitoring: review and perspectives. *Bioscience* **69**: 15–25.

Thirgood, S.J., Leckie, F.M. & Redpath, S.M. 1995. Diurnal and seasonal variation in line transect counts of moorland passerines. *Bird Study* **42**: 257–259.

Thomas, A., Speldewinde, P., Roberts, J.D., Burbidge, A.H. & Comer, S. 2020. If a bird calls, will we detect it? Factors that can influence the detectability of calls on automated recording units in field conditions. *Emu* **120**: 239–248.

Vold, S.T., Handel, C.M. & McNew, L.B. 2017. Comparison of acoustic recorders and field observers for monitoring tundra bird communities. *Wildl. Soc. Bull.* **41**: 566–576.

Whytock, R.C. & Christie, J. 2017. Solo: an open source, customizable and inexpensive audio recorder for bioacoustic research. *Methods Ecol. Evol.* **8**: 308–312.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **14**: 763–773.

Zilli, D., Parson, O., Merrett, G.V. & Rogers, A. 2014. A hidden Markov model-based acoustic cicada detector for crowdsourced smartphone biodiversity monitoring. *J. Artif. Intell. Res.* **51**: 805–827.

Received 7 May 2024;
Revision 11 October 2024;
revision accepted 16 October 2024.
Associate Editor: Juan Navedo

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Example Oystercatcher vocalization.

Figure S2. Example Lapwing vocalization.

Figure S3. Example Golden Plover vocalization.

Figure S4. Example Curlew vocalization.

Figure S5. Comparison of the outcomes of the acoustic and field surveys for Oystercatchers.

Figure S6. Comparison of the outcomes of the acoustic and field surveys for Lapwings.

Figure S7. Comparison of the outcomes of the acoustic and field surveys for Golden Plovers.

Figure S8. Comparison of the outcomes of the acoustic and field surveys for Curlews.