Ecological Informatics 55 (2020) 101014

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Using visualization and machine learning methods to monitor low detectability species—The least bittern as a case study

Elizabeth Znidersic[a,b,*], Michael Towsey[c], W.K. Roy[b], Sarah E. Darling[d], Anthony Truskinger[c], Paul Roe[c], David M. Watson[a]

[a] Institute for Land, Water and Society, Charles Sturt University, Elizabeth Mitchell Drive, Albury 2640, NSW, Australia
[b] Oak Ridge National Laboratory, Environmental Sciences Division, Oak Ridge, TN 37831, USA
[c] QUT Ecoacoustics Research Group, Science and Engineering Faculty, Queensland University of Technology, Australia
[d] Oak Ridge National Laboratory, Facilities and Operations Directorate, Oak Ridge, TN 37831, USA

ARTICLE INFO

Keywords:
Least Bittern
Bird call recognition
Conservation technology
Ecoacoustics
Acoustic ecology
Machine learning

ABSTRACT

Long duration acoustic monitoring is becoming an increasingly popular approach to extend survey effort by using autonomous sensors to passively collect data over large temporal and spatial scales. This is of particular benefit when attempting to detect a species whose temporal vocalization strategy is unknown, and whose small population size reduces detection probability. It is also of benefit in environments that are logistically difficult to access such as wetlands. We investigated the vocalization strategy of the Least Bittern (Ixobrychus exilis), a species of high conservation concern in the Western hemisphere and 'in need of management' in multiple states of the USA. The Least Bittern is a secretive marsh bird that is primarily detected by its vocalizations and call-playback surveys are typically used for population monitoring. To minimize disturbance to both the birds and their habitat, we deployed autonomous acoustic recording units and collected continuous 24-hour audio recordings for 30 days. The resultant accumulation of data necessitated an automated method to assist with analysis and interpretation. We successfully applied a novel soundscape technique—long-duration, false-color (LDFC) spectrograms—to visually confirm presence of Least Bittern from the 'coo coo coo' vocalization associated with breeding. In addition, we used a machine learning technique to automate the acoustic event detection process. Peak vocalization times were then predicted from an annotated dataset of actual calls and subsequently used to develop an optimal acoustic survey strategy. The results of this research demonstrate how machine learning methods can search large data sets for a specific species. This information can then be used to optimize existing monitoring methods, to increase detection probability and to minimize associated costs.

1. Introduction

Monitoring is the baseline for all conservation management and generally it relies on population counts or presence/absence surveys. Monitoring methods must be adapted to best fit the species and the environments they inhabit. Wetland ecosystems are logistically complex to monitor. This is due to the usually damp and boggy substrate and the typically dense, yet fragile, vegetation structure. Wetlands worldwide are declining at a rate that exceeds many other ecosystems and the species they support are therefore, rapidly losing their habitat (Finlayson 2012; Hu et al. 2017). For some species, the challenges of monitoring in wetlands means insufficient data precluding development of objective monitoring protocols (Conway 2011). Without more information about occurrence patterns and population dynamics of wetland-dependent species, determining which sites and species to manage depends on educated guesswork rather than objective evidence. Novel monitoring methods may help to address this problem.

Most marsh bird species are detected primarily from their vocalizations, with survey reliability affected by surveyor experience, temporal calling behaviour, time of day, survey effort, weather, and environmental factors (Conway and Gibbs 2011). Call-playback is the most common monitoring method for marsh birds as the calling rate of many species is thought to increase significantly with call elicitation, increasing the likelihood of detection. There is however, variation in responses rates between species and detection accuracy may vary (Conway and Gibbs 2005). Therefore, the effectiveness of call-playback and the broader impacts on species are unknown (Watson et al. 2018).

Recent technological advances in monitoring equipment have

* Corresponding author at: Institute for Land, Water and Society, Charles Sturt University, Elizabeth Mitchell Drive, Albury 2640, NSW, Australia.
E-mail address: eznidersic@csu.edu.au (E. Znidersic).

https://doi.org/10.1016/j.ecoinf.2019.101014
Received 6 August 2019; Received in revised form 21 September 2019; Accepted 2 October 2019
Available online 03 November 2019
1574-9541/

allowed wildlife managers to extend survey effort both temporally and spatially without the cost of paying technicians for extended fieldwork. Long-duration acoustic monitoring using autonomous recording units is becoming increasingly popular to monitor terrestrial ecosystems (Farina and Gage 2017). The advantages of using autonomous recording units include: 1. objective collection of data (reduces observer bias), 2. the creation of a permanent record of data for further analysis, 3. minimal disturbance and impact on species and habitat, and 4. an increase in the probability of detecting a species that seldom calls. The disadvantage in most instances is the resultant accumulation of recordings (big data) that are impractical to comprehensively review with experienced surveyors. Thus, practitioners are left with a problem that necessitates a reliable automated process for analysis and interpretation.

Long-duration false-colour spectrograms offer a novel way to interpret soundscapes over an extended period (Towsey et al. 2018). Each spectrogram represents a 24-h period reflecting the ecological soundscape of the location. The soundscape can then be analyzed broadly (that is, at a low resolution of 1 min or longer) using a variety of indices such as the Acoustic Complexity Index (Pieretti et al. 2010) and the Temporal Entropy index (Sueur et al. 2008), which are treated as surrogates for ecosystem or species diversity. In addition, a long-duration false-colour spectrogram can be used as a visual tool to identify both broad taxonomic groups and individual species. For example, the chorusing behaviour of birds, insects and frogs may continue for many minutes and consequently leave very obvious colour traces in a long-duration false-colour spectrogram. Alternatively, one or few individuals of a single species can leave a colour trace where they are the only species calling in that frequency band. Published examples to date include bat species and the Lewin's Rail (*Lewinia pectoralis brachipus*) (Towsey et al. 2018).

Machine learning methods are commonly applied to large datasets to identify individual vocal species, but these methods have limitations. Because the content of environmental recordings is unconstrained, the creation of labelled data sets to generate recognizers is frequently a time-consuming and expensive enterprise. Even with a large dataset of calls, depending on the species and complexity of the vocalization and the variable background noise, recognizers can be prone to a high error rate (false negatives and false positives).

In this study, we summarize a novel application of acoustic analysis and visualization methods to monitor the Least Bittern (*Ixobrychus exilis*). Our study site is the Oak Ridge Reservation, Tennessee, USA. Least Bittern is listed as a species 'In Need of Management' in Tennessee and is identified as a species of high concern in the Western Hemisphere by the Waterbird Conservation for the Americas, 2010. Much of the challenge involved in managing this species is concerned with their monitoring and detection, as they vocalize infrequently and their wetland habitat is mostly inaccessible (Bystrak 1981).

Our study includes five components: (1) Visual recognition of Least Bittern calls in long-duration false-colour spectrograms; (2) Training a predictor using machine learning methods to identify Least Bittern calls; (3) Devising a protocol to scan 30 days of soundscape recording using the same predictor; (4) Devising a Least Bittern monitoring strategy based on the calling patterns revealed by the above methods; and (5) Evaluating differences in the detection results between our methods and traditional monitoring protocols. Having summarized our findings, we demonstrate the practical utility of using *post-hoc* analysis of long-duration false-colour spectrograms to optimize monitoring strategies for low-detectability species, maximising both efficiency and comparability of resultant data.

2. Methods

2.1. Study site

The Oak Ridge Reservation is located in Roane and Anderson counties, East Tennessee. The property spans 13,549 ha and is recognized as the largest contiguous protected land ownership in the southern Valley and Ridge Physiographic Province (Roy et al. 2014). Oak Ridge has an average temperature range of 12.6–25.8 °C in May and 17.4–29.7C° in June and average rainfall 123 mm in May and 113 mm in June (U.S. climate data, 2019). Approximately 5% of the property comprise of wetlands ranging in size from < 2 ha to 62 ha. (Salk 2006).

2.2. Study species

The Least Bittern (*Ixobrychus exilis*) is a cryptic marsh bird and one of the smallest species of the heron family (Ardeidae). The current distribution range of Least Bittern extends throughout much of North, Central and South America. The nominate subspecies *exilis* is restricted to North America (Poole et al. 2009) and is listed as being of high concern in the North American Waterbird Conservation Plan and a species of migratory conservation concern in the Northeast by the US Fish and Wildlife Service (Kushlan et al. 2002). Least Bittern is representative of many secretive marsh birds that vocalize infrequently therefore requiring a considerable survey effort to confirm presence and a much greater effort to confidently infer absence. The effectiveness of detecting Least Bittern during multi-species marsh bird surveys has been questioned (Jobin et al. 2011). The call of the Least Bittern outside of the breeding season is seldom heard (Brewster 1902). During the breeding season the soft 'coo coo coo' vocalization is heard, and is the primary call used in call-playback surveys (Conway 2011). Contradictory results from studies indicated an improvement in detection probability with call-playback (Gibbs and Melvin 1993; Swift et al. 1988), and others identifying no significant change or a reduction in detection probability with this method (Lor and Malecki 2002; Manci and Rusch 1988; Tozer et al. 2007). In 2011, a species-specific protocol for monitoring the Least Bittern was developed (Jobin et al. 2011) which eliminated the multiple species call-playback regime that is currently used in the Standardized North American Marsh Bird Monitoring Protocol (Conway 2011). This was due to concern over the efficacy of the short broadcast period of the call (30 s) to elicit a response during the breeding period (Bogner and Baldassarre 2002; Tozer et al. 2007) and the potential confounding effects of interspecific species call-playback regime.

2.3. Data collection

One Songmeter 3 (SM3) acoustic sensors (Kaleidoscope 2017) was deployed from 8 May 2017 to 6 June 2017. We recorded continuously at the Heritage Center, Greenway Powerhouse Trail (35° 55.868′ N, 84° 18.603′ W). The acoustic sensor was positioned ~80 cm above the ground over shallow water (~50 cm deep) and mounted on a metal stake. A Wildlife Acoustics program was pre-loaded to record 'continuously' (24 × 1-h WAVE files per day) in 16-bit stereo at a sampling rate of 24 kHz. Memory cards (SD) and D-cell batteries were changed at approximately 10-day intervals to ensure uninterrupted recording.

2.4. Data visualization using long-duration false-colour (LDFC) spectrograms

All recordings were divided into consecutive, non-overlapping, one-minute segments of audio, down-sampled to 22,050 samples per second. Each one-minute segment was converted to an amplitude spectrogram by calculating an FFT (with Hamming window) over non-overlapping frames (width = 512). Each spectrum of 256 values (bin width = ~43.1 Hz) was smoothed using a moving average filter (width = 3). Decibel spectrograms were prepared by converting Fourier coefficients (A) to decibels using $dB = 20 \times \log10(A)$. In addition to the amplitude and decibel spectrograms, a third noise-reduced spectrogram was prepared by subtracting the modal decibel value of each

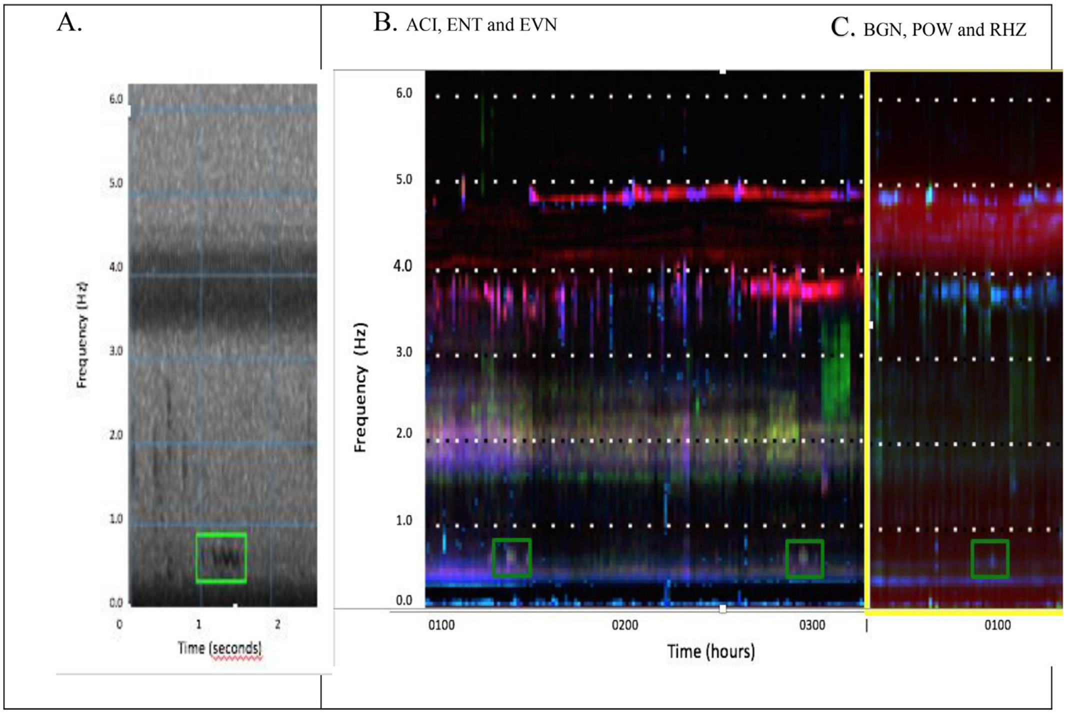

Fig. 1. The Least Bittern 'coo-coo-coo' vocalization (in green boxes) in A. a standard spectrogram, B. in a long-duration false-colour spectrogram (ACI, ENT and EVN indices) and C. in a long-duration false-colour spectrogram (BGN, POW and RHZ indices). Note the less obvious Least Bittern call in 2C due to the combination of indices. In both spectrograms the x-axis represents time and the y-axis is frequency. However, the scale is different in both cases. The x-axis ticks are at one-second intervals (A) and at one-hour intervals (B and C). The y-axis spans 6 kHz.

frequency bin from every value in the bin (Towsey 2017).

As a first step to preparing LDFC spectrograms, six *spectral acoustic indices* were calculated, one for each one-minute segment of recording. In this context, a *spectral index* is a 256-element vector, each element of which is derived from one frequency bin of a one-minute spectrogram. Each index can be understood as a mathematical function that summarizes some aspect of the distribution of acoustic energy in the frequency bin from which it is derived (Towsey et al. 2014). These indices have been previously described, so are briefly summarized here:

1. Acoustic Complexity Index (ACI): Calculated using the method of Pieretti et al. (2010). It measures relative change in acoustic intensity (A) in each frequency bin, f, of the amplitude spectrogram:

$$\mathrm{ACI}[\mathrm{f}] = \sum_{\mathrm{i}} |A_{\mathrm{if}} - A_{\mathrm{i-1,f}}| / \sum_{\mathrm{i}} A_{\mathrm{i}}$$

where i is an index over all frames and f is an index over all frequency bins.

2. Temporal Entropy index (ENT): (Sueur et al. 2008) A measure of the dispersal of acoustic energy through the frames of each frequency bin. The squared amplitude values in each frequency bin are normalized to unit area and treated as a probability mass function (*pmf*). The entropy of the *pmf* vector is a measure of its energy 'dispersal' through time and is calculated as:

$$\mathrm{H_t}[\mathrm{f}] = \sum_{\mathrm{i}} \log_2(\mathrm{pmf_{if}})/\log_2(N)$$

where i is an index over frames and N is the number of frames. To obtain a more 'intuitive' index, we convert $\mathrm{H_t}[\mathrm{f}]$ to 'energy concentration':

$$\mathrm{ENT}[\mathrm{f}] = 1 - \mathrm{H_t}[\mathrm{f}]$$

3. Event Count index (EVN): (Towsey 2017) A measure of the number of acoustic events per second in each frequency bin, f, of the noise-reduced decibel spectrogram. An event is counted each time the bin's (noise-reduced) decibel value crosses the 3-dB threshold from below.
4. Background Noise index (BGN): The modal decibel value in each frequency bin of the decibel spectrogram (Towsey 2017).
5. Power index (POW): The maximum decibel value in each frequency bin of the noise-reduced decibel spectrogram.
6. Horizontal Ridge Count index (RHZ): Also derived from the noise-reduced decibel spectrogram. Many bird songs consist of whistles with harmonics which appear in the standard grey-scale spectrogram as horizontal ridges. These can be detected using a 5 × 5 ridge mask. Each element of RHZ is the average decibel value of ridge cells identified within the corresponding frequency bin, f. RHZ can be helpful for detecting bird call activity.

Recordings were processed into LDFC spectrogram images, one for each 24-h period (midnight to midnight). Each spectrogram was prepared by combining three spectral acoustic indices, assigning one index to each of the red, green and blue colour channels using the method described in Towsey et al. (2014). We assigned ACI, ENT and EVN to the red, green and blue channels respectively to produce one LDFC spectrogram and the BGN, POW and RHZ indices to produce a second LDFC spectrogram (Towsey 2017). These combinations were chosen because false-colour spectrograms are most informative when the three indices are minimally correlated. Because the acoustic indices are calculated at one-minute resolution, the resulting LDFC spectrograms are 1440 pixels wide (1440 min in 24-h) and 256 pixels deep (equivalent to the number of frequency bins given a frame size of 512).

2.5. Call prediction using a random forest regression model

Building a call predictor for the Least Bittern involved three steps: (1) preparing a data set; (2) training a machine learning model; and (3) testing its performance. An important departure from the standard methodology for building call recognizers is that we did not attempt to recognize individual calls (see Fig. 1.A) for an example spectrogram of a single call). Instead, working at one-minute resolution, we trained a Random Forest model (Breiman 2001) in regression mode to predict the number of Least Bittern calls in any one-minute segment. Because we cast the prediction problem as a regression task rather than a classification task, we refer to the trained model in this study as a call *predictor* rather than a call *recogniser*, to highlight the distinction. This approach greatly reduced the time involved in preparing annotated datasets compared to cutting out numerous individual calls and selecting appropriate negative instances. In addition, it allowed us to use spectral indices (already calculated at one-minute resolution to prepare LDFC spectrograms) as acoustic features. It is important to note that although there is only an integer number of calls per minute, the trained model outputs a real value for each input vector.

Two complete days of recording were annotated by counting the number of Least Bittern calls per minute. The first day (1 June 2017) was used for training purposes, and a second day (4 June 2017) was used for testing. Consequently, there were 1440 instances in each of the training and test sets. All six spectral indices were used as acoustic features. The dominant components of the Least Bittern call lie between 490 and 800 Hz. We therefore only used that portion of each spectral index lying in this frequency band which, given our sampling rate, was frequency bins 12–21 (10 bins inclusive). A single call consists of three to four syllables, having total duration of 0.6–0.7 s.

The Random Forest model was trained in regression mode using the WEKA Machine Learning Package (Frank et al. 2016). Multiple runs of 10-fold cross-validation were used to determine the optimum combination of acoustic features. The labelled data from the 4 June 2017 was used to test performance of the best trained predictor. Because this was a regression task (in which the prediction output is a real value as opposed to a predicted class or category), prediction accuracy was calculated as a correlation coefficient (between actual and predicted values) and prediction error as mean absolute error.

2.6. Establishing a protocol for use of the regression predictor on real data

The full 30 days of recording were divided into one-minute segments and the Random Forest prediction was assigned to each minute. Given that the prediction is a real value representing number of calls per minute, a threshold is required to decide whether or not the minute of recording should be audited for the presence of Least Bittern. With no prior application of this approach to use as a guide, we listened to minutes in order of their prediction score.

3. Results

3.1. Visualization of least Bittern vocalizations in LDFC spectrograms

Thirty days (720 h) of continuous acoustic recording were collected from one site. Activity in the 500–800 Hz band was located using the LDFC spectrograms and the presence of one or more Least Bittern calls was confirmed using standard-scale spectrograms (Fig. 1A). With some practice it was possible to achieve visual detection of the Least Bittern vocalization signature in LDFC spectrograms (Fig. 1B). Two LDFC spectrograms (using different combinations of spectral indices) were prepared for each day of recording (Fig. 2). Least Bittern vocalizations were more easily identified in the ACI, ENT and EVN LDFC spectrograms (Fig. 1B) than the BGN, POW and RHZ combination (Fig. 1C). Fig. 2 illustrates the different components of a 24-h soundscape, including: (1) the morning chorus at 05:30–06:00; (2) an evening chorus around 20:30–21:00; (3) cricket choruses represented by approximately horizontal red lines whose frequency rises in the morning and drops in the evening due to temperature sensitivity of cricket calls; and (4) rainfall indicated by vertical red band at 17:30. Least Bittern calls can be seen at 600 Hz between 05:00 and 06:00 in the top LDFC spectrogram (Fig. 2). The first vocalization was detected 21 May 2017 by reviewing visually the LDFC spectrograms and the last was detected 6 June 2017 when the acoustic sensor was retrieved.

3.2. Performance of the random forest regression model

Actual Least Bittern calls were counted for every minute of the 1 June 2017 recordings (Fig. 3A) and the 4 June 2017 recordings (Fig. 3B). On 1 June 2017, Least Bittern vocalized in 215 of the 1440 min (~15% of the day) and on 4 June 2017 in 317 min (~22% of the day) (Fig. 3A and 3B). The peak vocalizing times were sunrise (05:00–06:00) and sunset (20:00–20:45). The maximum number of calls in any minute was 14.

A training set of 1440 instances derived from 1 June 2017 included 215 positive instances (target calls/min > 0) and 1225 negative instances (target calls/min = 0). The test data set consisted of 317 instances where target calls/min > 0.

The above dataset was used for training and testing the Random Forest model to predict the number of calls per minute. Using 10-fold cross-validation, the optimum combination of acoustic features was found to include only three acoustic indices (ACI, ENT, EVN) and only frequency bins 12 to 18 inclusive, that is, from 470 to 775 Hz. The correlation between the predicted calls per minute and the actual calls per minute was 0.90 with a mean absolute error of 0.47.

The actual and predicted number of calls per minute on the test day (4 June) are shown in Fig. 3B and C. The cross-correlation was reduced to 0.64 because the trained model consistently under-predicted calls per minute where the actual values were greater than eight. The predictor worked more efficiently during the night-time when the acoustic space was relatively uncluttered but failed to predict correctly those calls from 07:00–08:30. This 90 min coincided with calling activity of the Mourning Dove (*Zenaida macroura*) and Green Frog (*Lithobates clamitans*), both of whose calls overlap the frequency band of the Least Bittern call.

3.3. Using the call predictor over long-term recordings

To simulate how the Least Bittern call predictor would be used with very long-duration recordings, all 30 days of recording were divided into one-minute segments and the Random Forest prediction (calls per minute) assigned to each minute. We decided on a protocol which selected five one-minute segments corresponding to the five highest prediction scores in each day for the 30 days (yielding a total of 150 min of recording) and then determined the true number of calls for each of the selected minutes (Fig. 4). In practice, the selected minutes (150 in this case) could be ranked by highest prediction score and the true number of calls per minute confirmed up to some chosen level of effort. As an example, using the data in Fig. 4, if a threshold of seven predicted calls per minute is set, 92% of the selected minutes contain Least Bittern calls at a cost of nine missed minutes (recall of 86%). If the threshold is then dropped to six predicted calls per minute, an additional three minutes containing Least Bittern calls would be retrieved at the cost of auditing 14 min that do not contain Least Bittern calls.

The call prediction sequence reveals information about the daily cycle of call frequency (counts per minute) of the target species. This in turn informs the optimal time to conduct a Least Bittern survey. The prediction scores for the first 14 days of recording (which did not contain a Least Bittern call) were averaged minute by minute. This provides a ‘baseline’ response of the predictor through the 24-h soundscape in the absence of Least Bittern calls. Likewise, the remaining 16 days of predictions (most of which contained Least Bittern calls at some time during the day) were averaged minute by minute. On the assumption that the statistics of the non-bittern soundscape predictions remain stable over the 30 days of recording, subtracting the baseline predictions (initial 14-day averages) from the 16-day averages should indicate the contribution of Least Bittern calls to the daily soundscape. Fig. 5 demonstrates that the peak vocalization times of the

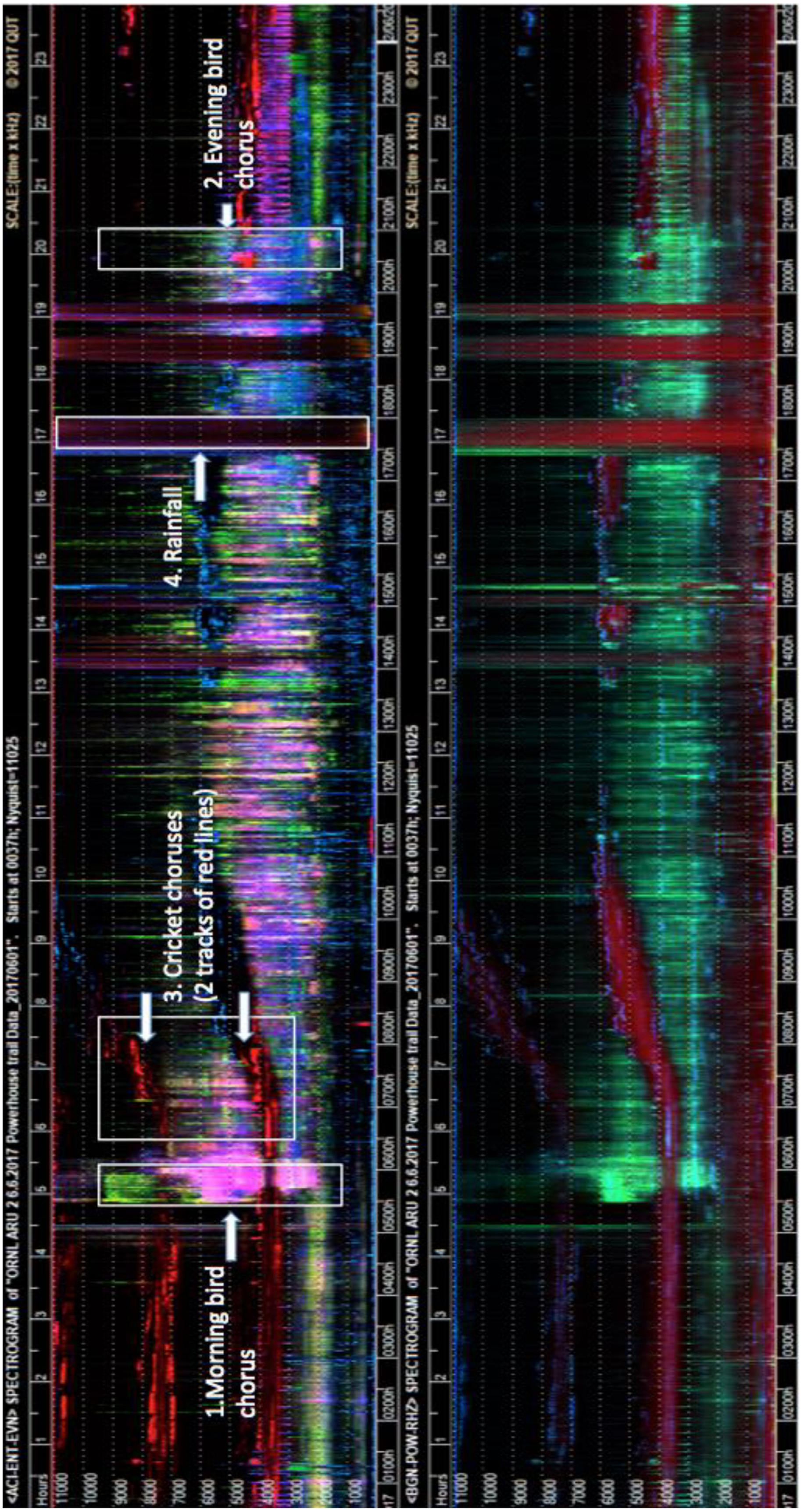

Fig. 2. Long-duration false-colour spectrogram from The Oak Ridge Reservation on 1 June 2017. The top image using ACI, ENT and EVN spectral indices, the bottom image using BGN, POW and RHZ spectral indices. The x-axis represents 24 hours (1 day) from 00:00 midnight to 11:59, and the y-axis frequency (Hz). Different components of the soundscape are illustrated including: (1) the morning bird chorus at 05:30–06:00 h; (2) the evening bird chorus around 20:30–21:00 hr; (3) cricket choruses represented two bands of horizontal red lines whose frequency rises in the morning due to temperature sensitivity of cricket calls; and (4) rainfall indicated by vertical red band at 17:30 hr.

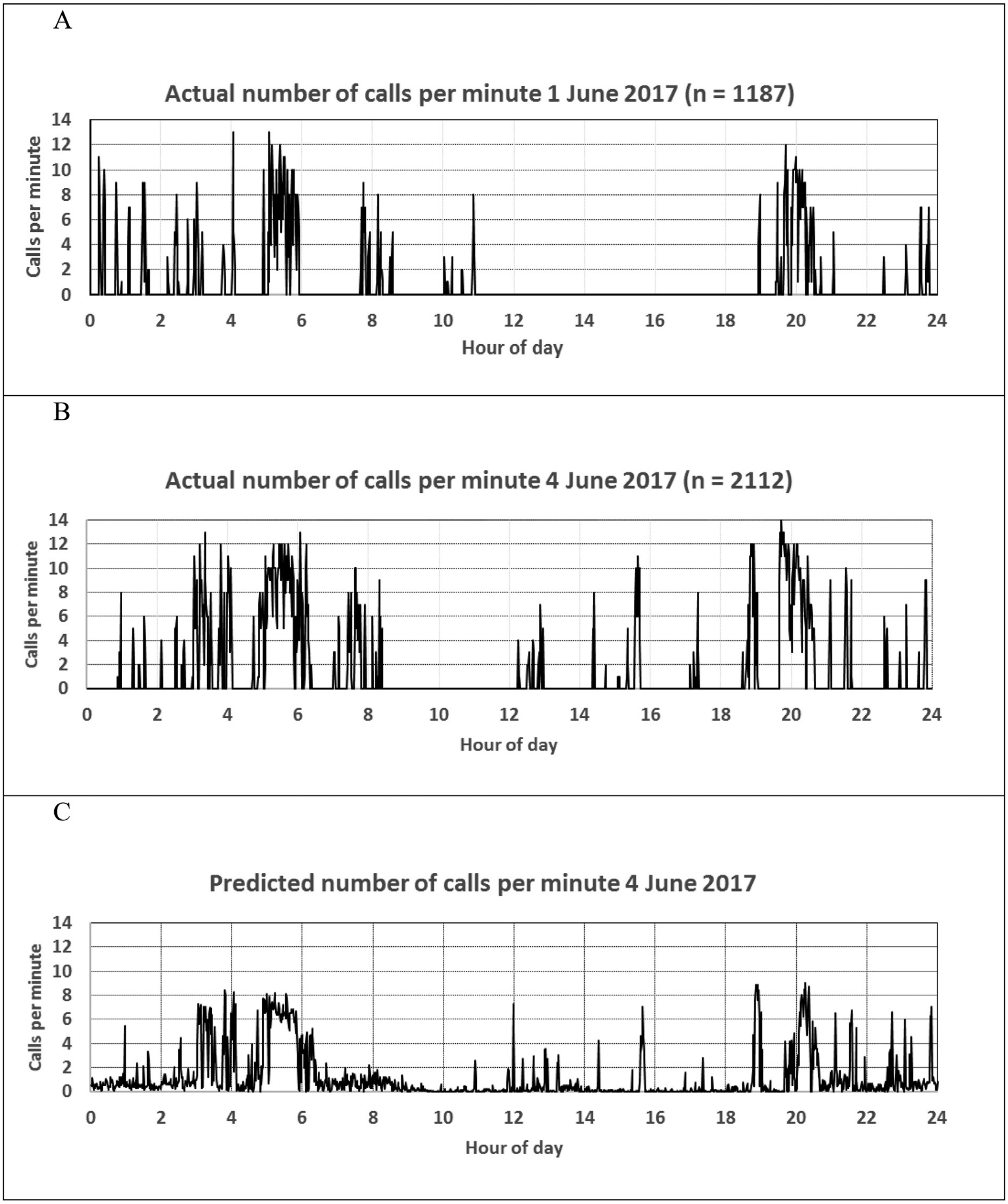

Fig. 3. The numbers of Least Bittern vocalizations per minute over a 24-h period (1440 min), midnight to midnight. A: The actual number of calls per minute for 1 June 2017. Total number of calls for the day is 1187. B: The actual number of calls per minute for 4 June 2017. C: The predicted number of calls per minute using a Random Forest recognizer.

Least Bittern calls is around dawn (04:30–05:30 h) and dusk (19:40–20:40 h).

4. Discussion

Species monitoring is the foundation of conservation management. In response to rapidly changing environmental and economic pressures, ecologists are adopting novel technologies to help them monitor faster and cost-effectively. In particular, researchers with widely different biological and computational skill sets now work collaboratively in their complementary areas of expertise. This may include experienced field technicians with local knowledge to deploy the acoustic sensors, skilled ornithologists to identify bird calls, computer scientists for data management and machine learning and ecologists to interpret the significance of the results. Given the requirement for such a mix of disciplines, it is important that the participants understand the limitations of each of the techniques being used.

This study investigates the calling behaviour of a cryptic marsh bird,

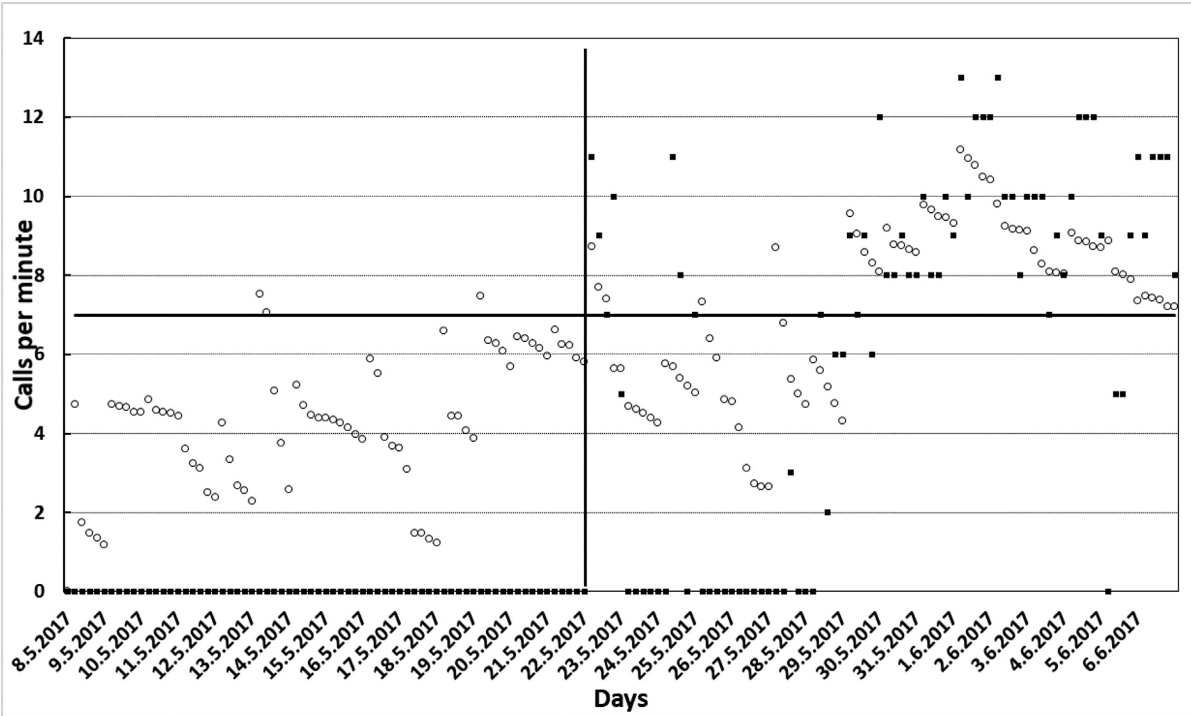

Fig. 4. Acoustic sensor deployment from 8 May 2017 to 6 June 2017, at Oak Ridge Reservation, Tennessee, USA. X-axis indicates the date, y-axis indicates the number of Least Bittern calls per minute. The hollow circles indicate the top five predictions for each of the 30 days. The solid squares indicate the true number of calls in each of the selected minutes. Note that the first actual bittern call did not occur until 21 May 2017 (vertical black line). At a threshold of seven calls per minute (the solid horizontal line), 47 of the 51 selected minutes (92%) contain actual Least Bittern calls.

the Least Bittern. We utilize 30 days (720 h) of continuous acoustic recording. A continuous field survey conducted over the same number of days by a field technician would be logistically and economically infeasible (Thomas and Marques 2012). However, it is also impractical to aurally sample 720 h of continuous recording. The use of automated machine learning methods was essential, but these present their own set of difficulties, in particular, the preparation of labelled data sets with which to train a mathematical model of the call of interest. Our approach was novel in a number of respects: 1. We prepared LDFC spectrograms which offer detailed visualization of the 30-day soundscape. An important advantage of this technique is that the soundscape becomes 'visible' to the ecologist without the requirement to train machine-learned models. After some practice, the calling sequences of the Least Bittern could be recognized in LDFC spectrograms; 2. We trained a Random Forest model to detect Least Bittern calling sequences using the same acoustic features that were used to prepare the LDFC spectrograms; and 3. We trained the Random Forest model in regression mode to predict calls per minute rather than individual calls. This greatly simplified the preparation of the labelled data sets. The model predictions were then used to inform a subsequent monitoring strategy.

4.1. Monitoring implications

The most commonly used monitoring methodology to detect wetland birds is call-playback survey (Conway 2011; Jobin et al. 2011). Call-playback guidelines from the North American Marsh Bird Protocol (Conway 2011) recommend the optimal survey time for Least Bittern as being the two hours around sunrise and sunset (Conway et al. 2004). In addition, the Marsh Monitoring Program 1995–2004 (Crewe et al. 2006) and Tozer et al. (2007) recommended call-playback surveys occur between 18:00 and sunset. We identified two peak vocalization times—between 04:30 to 05:30 h and 19:40 to 20:40 h (sunrise Tennessee in June, ~06:21 h and sunset ~20:55 h) (Fig. 5). Concerning optimal time of day to monitor, our results are congruent with these recommendations. The protocols also recommend a survey duration at any one survey point to be between 10 and 15 min, dependent on how many species are included in the call-playback protocol. Our results (actual and predicted vocalizations) suggest that even within the peak calling times, Least Bittern vocalizations are sporadic and a survey window of 10 to15 minutes might still lead to false negative conclusions. Obviously, the chance of false negative conclusions can be decreased by increasing survey time, but further analysis is required to

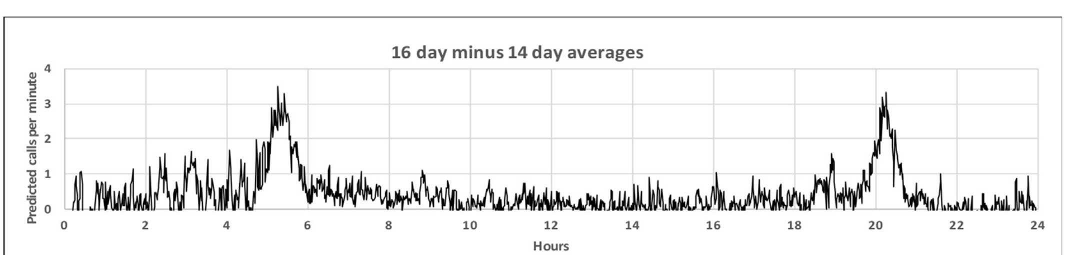

Fig. 5. Average predictions of Least Bittern calls per minute through a 24-h cycle. The predictions indicate the peak vocalization times are between 04:30–05:30 h and 19:40–20:40 h.

determine the optimal point-survey time.

In addition, current marsh bird monitoring protocols recommend up to three site visits (Conway 2011; Gibbs and Melvin 1993) during the breeding season, which usually spans approximately three months. Our study implies that if a Least Bittern survey took place during the first 14 days of deployment (8 May 2017 to 21 May 2017), an incorrect false negative conclusion would have been made. This suggests that more than one visit per month is required to confidently infer absence of Least Bittern. It must of course be noted that, in this study, an absence of calls prior to 22 May 2017 at one location does not mean absence of the species from the Oak Ridge Reservation. Least Bittern may have been present but not using the breeding 'coo-coo-coo' vocalization, or the birds may have been frequenting another location.

Failure to detect a species in an occupied habitat patch is a common sampling problem, particularly when the population is small, individuals are difficult to sample, and/or sampling effort is limited (Gu & Swihart 2003). Increasing survey effort by increasing field hours is likely to be costly but sample completeness is critical for presence-absence data and survey comparability (Watson 2017). Although some datasets may be adjusted using statistical models, the results can still be confounded by a lack of true absences, even when these models include pseudo-absence data (Stokland et al., 2011). The ability to make long-duration recordings and to detect a species of interest in LDFC spectrograms is very significant in terms of monitoring efficacy and cost.

4.2. Soundscape visualization

Wildlife practitioners are today monitoring with acoustic sensors over larger spatiotemporal scales than previously attainable. A large-scale approach is advantageous when monitoring a cryptic species such as the Least Bittern, but it does demand an automated approach to recording analysis. In this study, two computational techniques were applied to 720 h of acoustic data: LDFC spectrograms and automated call prediction. These two analytical tools answered different questions. By visualizing 30 days of soundscapes using LDFC spectrograms, we were able to confirm presence of the Least Bittern. It should be noted that some learning is required for effective interpretation of LDFC spectrograms. The wildlife practitioner must have a broad appreciation of the soundscape variability at the recording site, which in turn requires a prior understanding of the common biological and non-biological sound sources that contribute to the spectrogram. Only when the major features in an LDFC spectrogram and their variability are understood, should attention be turned to the less obvious features that may reveal a cryptic species of interest. Indeed, in our experience, attention to LDFC details has revealed cryptic species that were not the intended target species.

It is true that standard scale spectrograms (displaying approximately one-minute of audio in a desktop computer monitor) could also be scanned for presence of Least Bittern calls, but in our case, this would have meant reviewing 43,200 images (1440 min per day across 30 days) compared to 30 LDFC spectrogram images, each requiring about one minute of review.

Given that continuous recording over multiple weeks can still be a costly exercise depending on logistics, it should be noted that LDFC spectrograms can still be an effective visualization even when one does not record continuously. For example, we have prepared LDFC spectrograms from night-time only recordings when monitoring nocturnal owls, bats and frogs. Where it is appropriate, recordings can be limited to the hours around dawn and/or dusk. We find that about two to three hours of continuous recording is sufficient to locate a species in its soundscape.

4.3. Machine learning

The usual approach to bird call detection in acoustic recordings is to build a call recognizer for a binary classification task (call-of-interest versus not call-of-interest) using a set of typical positive and negative instances. What constitutes 'typical' is a significant issue in the training of an effective recognizer. The Least Bittern call is only about one-half second in duration but varied in duration from 0.3 to 0.7 s. Selecting an appropriate range of calls and determining appropriate start and end points for each of them would have been a time-consuming exercise. We observed that when this species started to call, it would often continue calling at the rate of one call every one to two second for as long as 30 s. This observation suggested that casting the recognition task as a regression task (that is, predicting calls per minute) would greatly simplify data set preparation since it was comparatively easy to count the number of calls per minute. Furthermore, the acoustic features had already been calculated at one-minute resolution in the preparation of the LDFC spectrograms.

Obviously, this approach can only work for species which are likely to call multiple times in a minute. Another disadvantage is that having feature vectors extracted over one-minute samples greatly increases the probability of confounding calls from other species in the same frequency band. In our case the Random Forest predictor failed to detect Least Bittern calls on the test day (6 June 2017) between 07:00 and 08:30 (Fig. 3C) when the Mourning Dove and Green Frog were vocally active. Despite these limitations, the method was effective in revealing the temporal vocalization pattern of the Least Bittern. Furthermore, the above method could be used as a filter, where selected minutes are then passed to a call recogniser trained in the usual way on a dataset of individual calls.

This study is the second to describe the use of LDFC spectrograms to identify a cryptic marsh bird species. The first concerned the Lewin's Rail, where its breeding call could be detected visually in an LDFC spectrogram (Towsey et al. 2018). Although LDFC spectrograms do not reveal the same temporal detail as standard spectrograms (whose resolution is typically about 20–40 ms per frame or pixel column), a visual identification of the call is possible if the target species is the dominant sound source in its frequency band and if there are no other confounding sounds in the same one-minute time period.

4.4. Conservation significance

As demonstrated in the Lewin's Rail study by Towsey et al. (2018), the visibility of Least Bittern vocalizations in LDFC spectrograms suggested that the spectral indices themselves could be used to train the Random Forest predictor. The Least Bittern call, like the Lewin's Rail call, occupies a relatively narrow bandwidth (400 to 800 Hz), below the typical range of most passerine calls. Where the species of interest calls in a crowded part of the sound spectrum (many passerines call in the 1 to 8 kHz band), this technique may not be so effective.

Despite this limitation, this study presents the first likely breeding record on the Oak Ridge Reservation of the Federally Imperiled Least Bittern (Natureserve 2018) from confirmation of the breeding call ('coo-coo-coo'). Just three incidental visual sightings have been made at this reservation from an extensive bird monitoring dataset that spans 67 years (Roy et al. 2014). Long-duration acoustic monitoring can be cost effective and reduce the impacts of disturbance on both species and habitat. In addition, the availability of a technique to visualize soundscapes greatly facilitates the processing of large amounts of acoustic data.

Acknowledgments

Thanks to the United States Department of Energy for their support on this project. Special thanks to Neil Giffen and Teresa Matthews who facilitated my visit to the Oak Ridge National Laboratory and for the marsh bird collaboration. In addition, thanks. to all staff at the Oak Ridge National Laboratory for sharing their knowledge and providing logistical support.

References

Bogner, H.E., Baldassarre, G.A., 2002. The effectiveness of call-response surveys for detecting least bitterns. J. Wildl. Manag. 66, 976–984.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Brewster, W., 1902. Voices of a new England Marsh. Bird-Lore 4, 43–58.
Bystrak, D., 1981. The North American breeding bird survey. In: Ralph, C.J., Scott, J.M. (Eds.), Estimating Numbers of Terrestrial Birds. 6. pp. 34–41 Cooper Ornithological Society. Studies in Avian Biology.
Conway, C.J., 2011. Standardized North American marsh bird monitoring protocol. Waterbirds 34, 319–346.
Conway, C.J., Gibbs, J.P., 2005. Effectiveness of call-broadcast surveys for monitoring marsh birds. Auk 122, 26–35.
Conway, C.J., Gibbs, J.P., 2011. Summary of intrinsic and extrinsic factors affecting detection probability of marsh birds. Wetlands 31, 403–411.
Conway, C.J., Sulzman, C., Raulston, B.A., 2004. Factors affecting detection probability of California Black Rails. J. Wildl. Mang. 68, 360–370.
Crewe, T., Timmermans, S., Jones, K., 2006. The Marsh Monitoring Program 1995–2004: A Decade of Marsh Monitoring in the Great Lakes Region. Bird studies Canada in Cooperation with the US Environmental Protection Agency and Environment Canada, Port Rowan, ON, Canada.
Farina, A., Gage, S. (Eds.), 2017. Ecoacoustics: The Ecological Role of Sounds. Wiley, New York.
Finlayson, C.M., 2012. Forty years of wetland conservation and wise use. Aquat. Conserv. Mar. Freshwat. Ecosyst. 22, 139–143.
Frank, E., Hall, M.A., Witten, I.H., 2016. The WEKA workbench. In: Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques”, 4th edition. Morgan Kaufmann. https://www.cs.waikato.ac.nz/ml/weka/book.html.
Gibbs, J.P., Melvin, S.M., 1993. Call-response surveys for monitoring breeding waterbirds. J. Wildl. Manag. 57, 27–34.
Gu, W., Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. Biological Conservation 116, 195–203.
Hu, S., Niu, Z., Chen, Y., Li, L., Zhang, H., 2017. Global wetlands: potential distribution, wetland loss, and status. Sci. Total Environ. 586, 319–327.
Jobin, B., Bazin, R., Maynard, L., McConnell, A., Stewart, J., 2011. Least bittern (*Ixobrychus exilis*) survey protocol. Waterbirds 34, 225–233.
Kaleidoscope, 2017. Wildlife Acoustics. 4.3.1 Documentation. Available at. https://www.wildlifeacoustics.com/images/documentation/Kaleidoscope.pdf, Accessed date: 19 January 2019.
Kushlan, J.A., Steinkamp, M.J., Parsons, K.C., Capp, J., Cruz, M.A., Coulter, M., Davidson, I., Dickson, L., Edelson, N., Elliot, R., Erwin, R.M., Hatch, S., Kress, S., Milko, R., Miller, S., Mills, K., Paul, R., Phillips, R., Saliva, J.E., Sydeman, B., Trapp, J., Wheeler, J., Wohl, K., 2002. Waterbird Conservation for the Americas: the North American Waterbird Conservation Plan, Version 1. Waterbird Conservation for the Americas, Washington, DC, USA, pp. 78.
Lor, S., Malecki, R.A., 2002. Call-response surveys to monitor marsh bird population trends. Wildl. Soc. Bull. 30, 1195–1201.
Manci, K.M., Rusch, D.H., 1988. Indices to distribution and abundance of some inconspicuous waterbirds on Horicon marsh. J. Field Ornithol. 59, 67–75.
Natureserve, 2018. NatureServe Explorer: An Online Encyclopedia of Life. Version 7.1. NatureServe, Arlington, Virginia Available at. http://explorer.natureserve.org, Accessed date: 20 January 2019.
Pieretti, N., Farina, A., Morri, D., 2010. A new methodology to infer the singing activity of an avian community: the acoustic complexity index (ACI). Ecol. Indic. 11, 868–873.
Poole, A. F., Lowther, P. E., Gibbs, J.P., Reid, F. A. &. Melvin, S. M. (2009). Least Bittern (*Ixobrychus exilis*), version 2.0. In The Birds of North America (A. F. Poole, Editor). Cornell lab of ornithology, Ithaca, NY, USA. Available at doi: https://doi.org/10.2173/bna.17.
Roy, W.K., Giffen, N.R., Wade, M.C., Haines, A.M., Evans, J.W., Jett, R.T., 2014. Oak Ridge Reservation Bird Records and Population Trends. ORNL/TM-2014/109. Oak Ridge National Laboratory, Oak Ridge, Tennessee.
Salk, M.S., 2006. Wetlands of the Oak Ridge Reservation. (Brochure).
Stokland, J.N., Halvorsen, R., Stoa, B., 2011. Species distribution modelling - Effect of design and sample size on pseudo-absence observations. Ecological Modelling 1800–1809.
Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid acoustic survey for biodiversity appraisal. PLoS One 3, e4065.
Swift, B.L., Orman, S.R., Ozard, S.W., 1988. Response of least bitterns to tape-recorded calls. Wilson Bull. 100, 496–499.
Thomas, L., Marques, T.A., 2012. Passive acoustic monitoring for estimating animal density. Acoust. Today 8, 35–44.
Towsey, M., 2017. The Calculation of Acoustic Indices Derived from Long-Duration Recordings of the Natural Environment. Available at. https://eprints.qut.edu.au/110634, Accessed date: 27 March 2018.
Towsey, M., Wimmer, J., Williamson, I., Roe, P., 2014. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecol. Informa. 21, 110–119.
Towsey, M., Znidersic, E., Broken-Brow, J., Indraswari, K., Watson, D.M., Phillips, Y., Truskinger, A., Roe, P., 2018. Long-duration, false-colour spectrograms for detecting species in large audio datasets. J. Ecoacoustics. 2, IUSWUI Available at. 10.22261/JEA.IUSWUI.
Tozer, D.C., Abraham, K.F., Nol, E., 2007. Short call-broadcasts fail to detect nesting least bitterns (*Ixobrychus exilis*). Northeast. Nat. 14, 637–642.
U.S. Climate Data, 2019. Oak Ridge Tennessee. https://www.usclimatedata.com/climate/oak-ridge/tennessee/united-states/ustn0370.
Watson, D.M., 2017. Sampling effort determination in bird surveys: do current norms meet best-practice recommendations? Wildl. Res. 44, 183–193. https://doi.org/10.1071/WR16226.
Watson, D.M., Znidersic, E., Craig, M., 2018. Ethical birding call playback and conservation. Conserv. Biol. https://doi.org/10.1111/cobi.13199.