Check for updates

Received: 22 January 2024 | Accepted: 22 November 2024

DOI: 10.1111/1365-2664.70000

BRITISH ECOLOGICAL SOCIETY Journal of Applied Ecology

PERSPECTIVE

Mitigating bias in long-term terrestrial ecoacoustic studies

David Jarrett[1,2] | Ross Barnett[3] | Tom Bradfer-Lawrence[4] | Jérémy S. P. Froidevaux[5,6] | Kieran Gibb[7] | Pauline Guinet[8] | Jack Greenhalgh[9] | Becky Heath[10,11] | Alison Johnston[12] | José Joaquín Lahoz Monfort[9] | Alex Rogers[13] | Stephen G. Willis[1] | Oliver Metcalf[14]

[1]Department of Biosciences, Durham University, Durham, UK; [2]BTO Scotland, Stirling, UK; [3]School of Biological and Environmental Sciences, University of Stirling, Stirling, UK; [4]RSPB Centre for Conservation Science, Edinburgh, UK; [5]Biological & Environmental Sciences, University of Stirling, Stirling, UK; [6]CESCO, Museum National d'Histoire Naturelle, CNRS, Sorbonne University, Paris, France; [7]School of Informatics, University of Sussex, Brighton, UK; [8]Muséum National d'Histoire Naturelle, Paris, France; [9]Instituto Pirenaico de Ecología, Jaca, Huesca, Spain; [10]Department of Zoology, University of Cambridge, Cambridge, UK; [11]Kings College Cambridge, Cambridge, UK; [12]Centre for Research into Ecological and Environmental Modelling, School of Mathematics and Statistics, University of St Andrews, St Andrews, UK; [13]Department of Computer Science, University of Oxford, Oxford, UK and [14]Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence
David Jarrett
Email: david.jarrett@bto.org

Funding information
Leverhulme Trust, Grant/Award Number: DS-2017-070 and ECF-2020-571; Scottish Alliance for Geoscience, Environment and Society; Company of Biologists, Grant/Award Number: B133; UK Acoustics Network, Grant/Award Number: EP/V007866/1

Handling Editor: Vitor Paiva

Abstract

1. Long-term biodiversity monitoring is needed to track progress towards ambitious global targets to reduce species loss and restore ecosystems. The recent development of cheap and robust acoustic recording devices offers a cost-effective means of gathering standardised long-term datasets.
2. Accounting for sources of bias in ecological monitoring and research is a fundamental part of the study design process. To highlight this issue in the context of long-term terrestrial ecoacoustic monitoring, here we collate and discuss sources of bias arising from (i) hardware devices, (ii) firmware, software and analysis tools and (iii) the deployment environment.
3. One important source of bias is unavoidable changes in recording hardware—to demonstrate how this potentially introduces bias, we present two case studies comparing the output from simultaneous recordings from different recorders.
4. To mitigate biases, we recommend effective documentation of environmental and hardware-related variables, as well as a long-term data storage strategy that facilitates reanalysis. Additionally, the use of regular calibration tests to measure variation in the acoustic detection space will facilitate analytical approaches or post-hoc AI solutions that remove unwanted biases.
5. *Synthesis and applications*: The sources of bias and suggested mitigations described here will be of relevance to hardware manufacturers, ecological researchers and conservation practitioners. Researchers and conservation practitioners must be fully aware of relevant biases when designing long-term ecoacoustic studies and should incorporate appropriate mitigations into their study design.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2025 The Author(s). *Journal of Applied Ecology* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

KEYWORDS
acoustic indices, bias, bioacoustics, biodiversity, ecoacoustics, monitoring, passive acoustics

1 | BACKGROUND

Meeting ambitious global targets to reverse species extinction rates and restore ecosystems (IPBES, 2019) requires informed policy choices. Robust, standardised, long-term monitoring of species, biological communities and ecosystems is needed to provide the data underpinning such decision-making (Pereira & Cooper, 2006). Traditional fieldworker-led monitoring at scales sufficient to generate long-term trends is labour-intensive and can be logistically challenging (Schmeller et al., 2009), limiting the spatial and temporal extent of these datasets. However, technological developments and the relatively low cost of modern acoustic recorders (Hill et al., 2019) present opportunities for the long-term monitoring of sound-producing animals and soundscapes at much greater spatial and temporal scales than is possible with human observers (Darras et al., 2019; Obrist et al., 2010; Roe et al., 2021; Sethi et al., 2021).

Long-term ecological studies are often hampered by sampling biases, and there is a deep literature highlighting these challenges and offering suggestions for their resolution (e.g. Boyd et al., 2022, 2023; Dobson et al., 2020; Magurran et al., 2010; Meng, 2018). Various papers discuss the benefits and challenges of passive acoustic monitoring (e.g. Darras et al., 2019; Gibb et al., 2018; Sugai et al., 2019), but to date, the implications of biases in the context of long-term monitoring with these technologies have not been fully examined. As such, in this paper, we highlight and address issues that arise specifically when using passive acoustic monitoring for long-term ecological studies, as this presents a unique set of challenges and potential biases.

Data from acoustic recorders is often assumed to be more consistent and reliable than data collected by human observers because many of the biases associated with the latter are removed (Darras et al., 2018, 2019). However, there are many factors that can introduce biases during data collection with acoustic recorders and analysis with acoustic indices or species recognisers (Hartig et al., 2024). Practitioners, researchers, software and hardware developers should be aware of where and how biases originate and have a strategy for mitigating these when designing or using acoustic methods for long-term monitoring programmes.

2 | SOURCES OF BIAS IN LONG-TERM ACOUSTIC PROJECTS

We consider 'bias' to be any factor that could introduce a systematic but unquantified alteration, distortion, or misrepresentation of acoustic data that might lead to inaccurate inference about long-term changes in ecosystems. These biases can emerge at any stage during data collection and analysis. Here, we divide these biases into three broad categories based on the source of the bias: hardware (Table 2), software (Table 3) and the deployment environment (Table 4). These biases will often alter variables such as the signal-to-noise ratio, sensitivity, polar pattern or frequency response (bold terms are explained in Table 1) which will lead to an alteration in the size of the acoustic detection space. Were a researcher unaware that a bias had altered the acoustic detection space in a long-term dataset, the researcher may incorrectly infer (or miss) variation in a species' occupancy, density, or behaviour. However, there are various measures that a researcher can take to (i) mitigate and minimise biases that cause variation in the acoustic detection space; (ii) test that the acoustic detection space remains constant; and (iii) account for variation in the acoustic detection space in the data analysis.

2.1 | Hardware

Changing the hardware used in long-term acoustic projects will influence the data collected (Adams et al., 2012; Luna-Naranjo et al., 2024). Audio recordings can be influenced by power sources (Miquel et al., 2022), processor chip architecture (Hayman et al., 2017) or the type, age and quality of microphone (Darras et al., 2020; Turgeon et al., 2017). Even using the same make and model of recorder is no guarantee of consistency as availability will change over time: manufacturers discontinue or improve products, components are superseded, and there can even be variation between batches of the same component. All of these changes potentially alter the target signal representation and thus the acoustic detection space.

2.2 | Analysis tools, firmware and software

Widely available recording analysis tools are regularly updated, affecting both acoustic indices (Sueur et al., 2008; Villanueva-Riviera & Pijanowski, 2018) and species recognisers (Center for Conservations Bioacoustics, 2023; Kahl et al., 2021). For example, the widely used BirdNET sound ID tool is periodically updated to cover a greater range of species and new vocalisations (Kahl et al., 2021). This means that later analyses must be run over the whole dataset, otherwise new outputs would not be directly comparable to older outputs.

Device firmware and driver updates potentially improve recorder reliability, recording quality, power usage and make other minor fixes, but there is little literature on how these patches may affect acoustic indices or species recognition algorithms (Open Acoustic Devices, 2024; Wildlife Acoustics, 2024). For example, gain settings changed in early AudioMoth firmware versions so that identical gain

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Glossary of key terms used in this paper.

Term	Definition	Explanation
Signal-to-noise ratio	The ratio of the amplitude of the target sound signal to the amplitude of background noise	This describes how well the target signal, for example an animal vocalisation, stands out against background noise. A higher ratio means a clearer target signal
Sensitivity	The efficiency of the microphone in turning acoustic energy into an electrical signal	This determines how faint a sound can be detected. Higher sensitivity is needed for detecting quieter sounds
Polar pattern	The directional sensitivity of a microphone, or how well it picks up sound from different angles	Polar patterns can be visualised in 2D or 3D space. Common polar patterns are cardioid, omnidirectional and bidirectional
Frequency response	The range of sound frequencies that a microphone can capture and its sensitivity within that range. It can be represented graphically with a response curve	A flat response microphone is equally sensitive to all frequencies, while a shaped response curve varies in its sensitivity to different frequencies
Detection space	The geometrical space that the recording device samples effectively. Detection space may also be referred to as detection range or radius	Determined by the amplitude, directionality and frequency of the target species' vocalisations, the ambient sound level and the microphone features: signal-to-noise ratio, polar pattern, sensitivity and frequency response

settings resulted in different recording levels between v1.2.0 and v1.4.0 (Lapp et al., 2023).

Choices made when processing acoustic data can affect acoustic indices patterns (Bradfer-Lawrence et al., 2023, 2024; Metcalf et al., 2024) and species detectability (Perea & Tena, 2020). Additionally, different software programs can produce different results despite theoretically running the same procedure. This is well documented with acoustic indices (Bradfer-Lawrence et al., 2024), where differences in default settings, analytical steps and even internal representations of audio files can all influence the final output.

2.3 | Deployment environment

Changes to the deployment environment will cause variation in the effective detection space (Darras et al., 2016). Anthropogenic changes to soundscapes (e.g. increased noise from roads) over time are likely to be a widespread phenomenon (Fairbrass et al., 2017), as is changing vegetation and climatic conditions (Haupert et al., 2023; Sánchez-Giraldo et al., 2020; Thomas et al., 2020). All of these factors could introduce systematic variation in sound propagation or attenuation and thus the signal-to-noise ratio (Lapp et al., 2023) although the relationship between land use, vegetation and effective detection space of recording units is complex (Darras et al., 2016). Damage to recording units by animals is also common and may alter the recorded signals. Damage might include obvious destruction by domestic livestock but can be more subtle, such as invertebrates building nests in recorder housings or consuming wind shields.

2.4 | Summary

Separating the effects of biases from genuine variation in target signals may be challenging and requires consideration from practitioners, researchers, and software and hardware developers. In the remainder of this paper, we collate and summarise sources of bias that may arise when using acoustic recording in long-term monitoring programmes (Tables 2–4). We highlight the project stages where these biases should be considered and propose potential mitigations. Potential mitigations and other important considerations are explored further in the main text. To illustrate how biases could influence ecological inference we present two real-world case studies.

3 | CASE STUDIES

To provide examples of some of the biases we outline in Tables 2–4, and how variation in recording performance can translate into differences in outputs, we present two case studies; one using acoustic indices to characterise soundscape patterns, and one using a species

TABLE 2 Sources of bias arising from the deployment environment.

Source of bias	How could this affect long-term acoustic studies?	At what stage of the project should the bias be considered?					Proposed approach	Who should address the bias?			
		Study design	Equipment purchase	Settings	Deployment	Analysis		Field-worker	Project planner	Software developer	Hardware manufacturer
Climate change	Climate change is causing non-random changes in weather patterns that could introduce systematic bias in sound attenuation	✓	✓	✓	✓	✓	Use windshields to minimise the influence of changing wind patterns on capturing acoustic signals. Wind noise reduction algorithms can be used post-hoc to remove (or add) wind noise to make recordings comparable	✓	✓		✓
Changes to habitat and human infrastructure	Long-term changes in habitat structure and human infrastructure could lead to non-random changes in the sound attenuation. Changes in ambient noise levels may also occur as a result of human activity	✓		✓	✓	✓	Monitoring protocol should include recording habitat and infrastructure covariates for each recorder location. Use calibration tests for periodic assessment of sound attenuation rates across different frequencies (Section 4.2.2)	✓	✓	✓	
Animal damage	Long-term variation in the abundance and distribution of animals that could interfere with, or damage, recorders may require changing monitoring locations, or alterations to recorder casing or attachment method	✓	✓		✓		Ensure equipment and casing is as robust as possible to potential damage (Section 4.1). Protocol should include data checking and cleaning to remove compromised recordings	✓	✓		

Note: The 'proposed approach' column includes measures to avoid, mitigate or quantify the bias.

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Sources of bias arising from software.

Source of bias	How could this affect long-term acoustic studies?	At what stage of the project should the bias be considered?					Proposed approach	Who should address the bias?			
		Study design	Equipment purchase	Settings	Deployment	Analysis		Field-worker	Project planner	Software developer	Hardware manufacturer
Software, classifiers, and training datasets	Updates to software (e.g. BirdNET or Kaleidoscope) or augmented training datasets may affect acoustic indices values or vocalisation detection			✓		✓	Document software version used, and regularly reanalyse full datasets. If impractical to retain all data, save a subset (e.g. 10%), to allow calibration between old and new recordings. Software developers need to maintain access to legacy versions		✓	✓	
Recorder firmware and drivers	Firmware or driver updates may alter recording parameters, which could influence acoustic recordings	✓		✓		✓	Report firmware version used and conduct simultaneous recordings to allow calibration between old and new recorder set-ups. Software developers could ensure access to legacy versions and report results of a standardised calibration test (see Section 4.2.2)		✓	✓	
Software and data processing choices	Internal representations of acoustic data may differ among packages and operating systems. Pre-processing and noise reduction algorithms can alter data patterns	✓		✓		✓	Always report versions of software and hardware used. Document all data processing choices, retain code in publicly accessible repositories		✓	✓	

Note: The 'proposed approach' column includes measures to avoid, mitigate or quantify the bias.

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Sources of bias arising from hardware.

Source of bias	How could this affect long-term acoustic studies?	At what stage of the project should the bias be considered? Study design	Equipment purchase	Settings	Deployment	Analysis	Proposed approach	Who should address the bias? Field-worker	Project planner	Software developer	Hardware manufacturer
Hardware availability	Components and devices may be discontinued. Changes in microphones, power sources, casings, and SD cards will affect the recording quality	✓	✓	✓	✓	✓	Use a standardised calibration test when components or recorders change, and simultaneous recordings with the old and new set-up will facilitate quantification of variation. Hardware manufacturers should report the results of standardised calibration tests when components change		✓		✓
Physical degradation of components	Equipment degradation could lead to deteriorating performance, giving a false suggestion of temporal change	✓	✓		✓	✓	Calibrate equipment regularly. During analyses consider the potential of signal degradation		✓		✓
Change in available settings	Gain settings affect amplitude but also increase the possibility of clipping and, to a lesser extent, recorder self-noise			✓			Device manufacturers should ensure that when available settings change, there is full documentation, and where possible the ability to use legacy settings. Categorical labels need to be defined with absolute values, for example 'Medium' gain is equivalent to a specific decibel value	✓	✓		✓
Variation in hardware quality between batches or versions	Device components can vary over time in both quality and availability	✓	✓	✓			Consider purchasing spares to cover the lifetime of the project. Where possible, calibrate recorders to ensure consistency. Device manufacturers should use calibration tests among batches to document variation	✓	✓		✓

Note: The 'proposed approach' column includes measures to avoid, mitigate or quantify the bias.

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

classifier to identify bird calls. Both case studies collected simultaneous recordings at the same locations with two different recorder models. These are intended to be illustrative of the types of differences that might arise based on choices made during hardware selection.

3.1 | Case study 1: Soundscapes in the Brazilian Amazon

We collected 24 h of continuous recordings in old-growth forest in the Brazilian Amazon using an FL-BAR and an AudioMoth attached to the same tree. Recorder type had a strong impact on what was recorded and the signal-to-noise ratio of recordings; for example, the AudioMoth detected insect sounds between 12 and 16 kHz not picked up by the FL-BAR, whilst the inverse was true at lower frequencies. For each 1-min file we calculated the Acoustic Complexity Index (ACI) and the Bioacoustic Index using the soundecology package (v1.3.3, Villanueva-Riviera & Pijanowski, 2018) in R (v.4.3.1, R Core Team, 2023). We compared the two time-series visually using general additive models produced with the geom_smooth function including standard error, in the R package ggplot2 (Wickham, 2016), with all other arguments left as default.

Rainfall drove clear periodicity in the acoustic indices values (particularly between 01:00 and 02:00: Figure 1). Bioacoustic Index values were minimally affected by recorder type; the AudioMoth values closely tracked those of the FL-BAR (Figure 1), albeit slightly lower on average throughout (mean difference −4.36 ± 3.86 SD). In contrast, the ACI values differed between the recorders. The AudioMoth recorded many loud raindrops as they hit the hard casing housing the microphone, resulting in high ACI values (Figure 1, Bradfer-Lawrence et al., 2019). The FL-BAR's microphone is positioned below the recorder housing; while less impacted during rainfall, the foam cover may absorb water afterwards (and waterlogged microphones may have a very small range until they dry out—which is often many hours later). This potentially underlies the temporal lag in ACI values (Figure 1). While the FL-BAR microphones generally have a better signal-to-noise ratio, at some points the AudioMoth appeared to have a higher signal-to-noise ratio, particularly at higher frequencies and after periods of rainfall. These patterns mirror a much wider-scale study by Zhang et al. (2024), who reported substantial variation in acoustic indices values from different recorders when the signal-to-noise ratio was low.

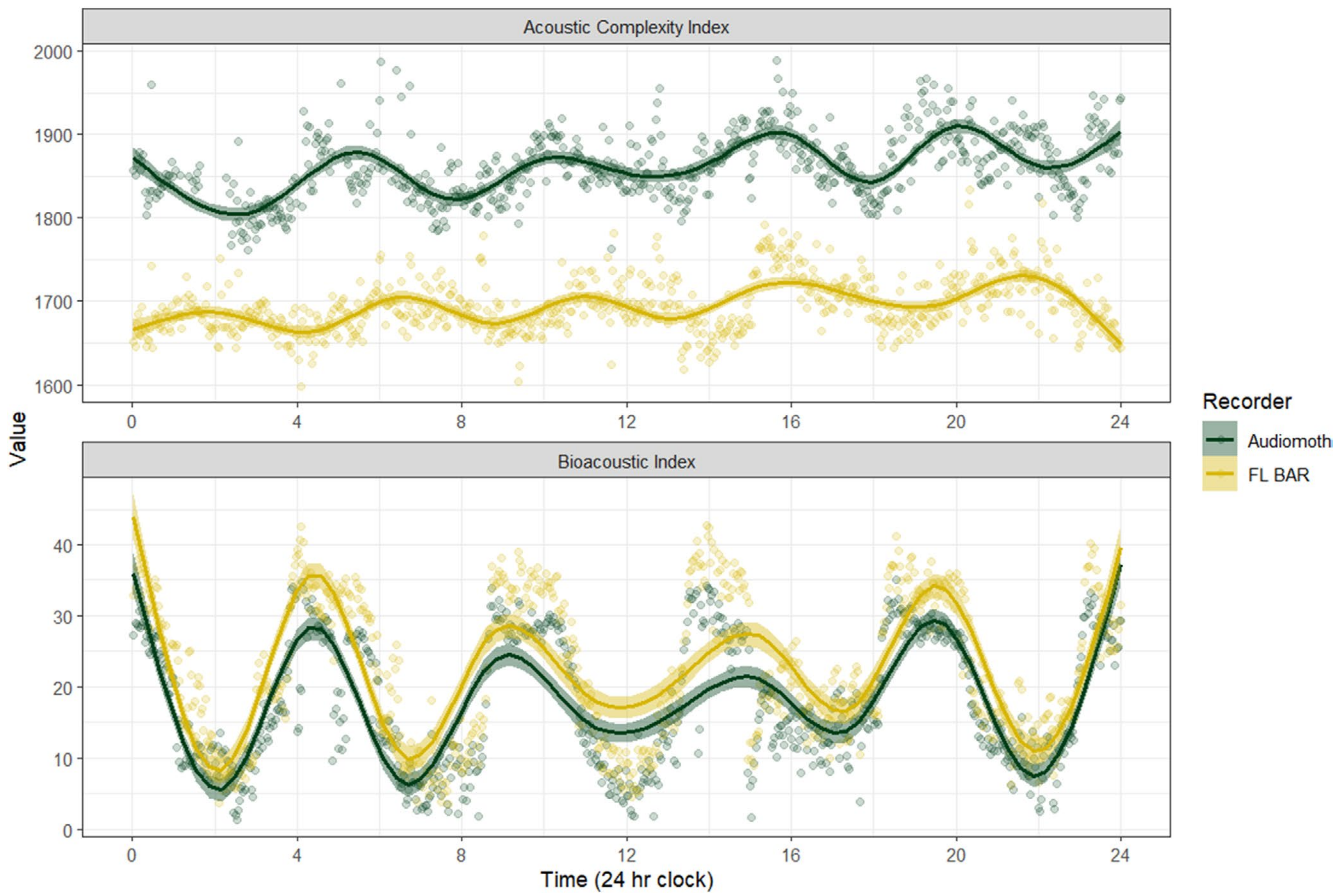

FIGURE 1 A comparison of Acoustic Complexity Index (top) and Bioacoustic Index (bottom) values derived from simultaneous recordings from an AudioMoth v1.2 and a Frontier Labs Bioacoustic Recorder. Points show values calculated in R using the soundecology package; the smooths are outputs from generalised additive models.

3.2 | Case study 2: Avian species in the Pyrenean temperate forest

We used an AudioMoth and a SongMeter Micro attached to the same tree in the Spanish Pyrenees to collect 5980 min of recordings over 22 days. We analysed the recordings using BirdNET V2.4 (Kahl et al., 2021) which returned 1081 and 1872 detections from the AudioMoth and SongMeter Micro datasets respectively. We did not manually vet each detection as we were not making ecological inferences. In line with Kahl and Wood (2024), we expected differences in the BirdNET scores between the two recorders.

Twenty-six species were detected with the AudioMoth and 23 with the SongMeter Micro. Twelve species (46%) were only recorded by the AudioMoth, and 9 species (39%) only by the SongMeter Micro; these included 4 and 3 species respectively that were likely false positives because they do not occur in this habitat (e.g. ducks and waders). The 9 species with the most detections were the same across both devices. The main findings are summarised in Figure 2; the number of detections differed in some cases (e.g. 447 detections of European robin *Erithacus rubecula* from the SongMeter Micro recorder vs. 169 from the AudioMoth) but was similar in other species (e.g. 256 vs. 272 for Short-toed Treecreeper *Certhia brachydactyla*). However, these similar scores disguise non-trivial variation in the confidence values obtained from both recorders. In a long-term monitoring program where vocalisation activity rate was used to estimate species density (Pérez-Granados et al., 2019) changing recorders in the middle of the monitoring period could compromise the data, and so a period of simultaneous deployment would be essential (see Section 4.2).

4 | IMPLICATIONS AND GUIDANCE

Biases can be dealt with in advance (i.e. prevention) or overcome after they have emerged (i.e. mitigation; Dobson et al., 2020). We identified 10 potential sources of bias in long-term acoustic

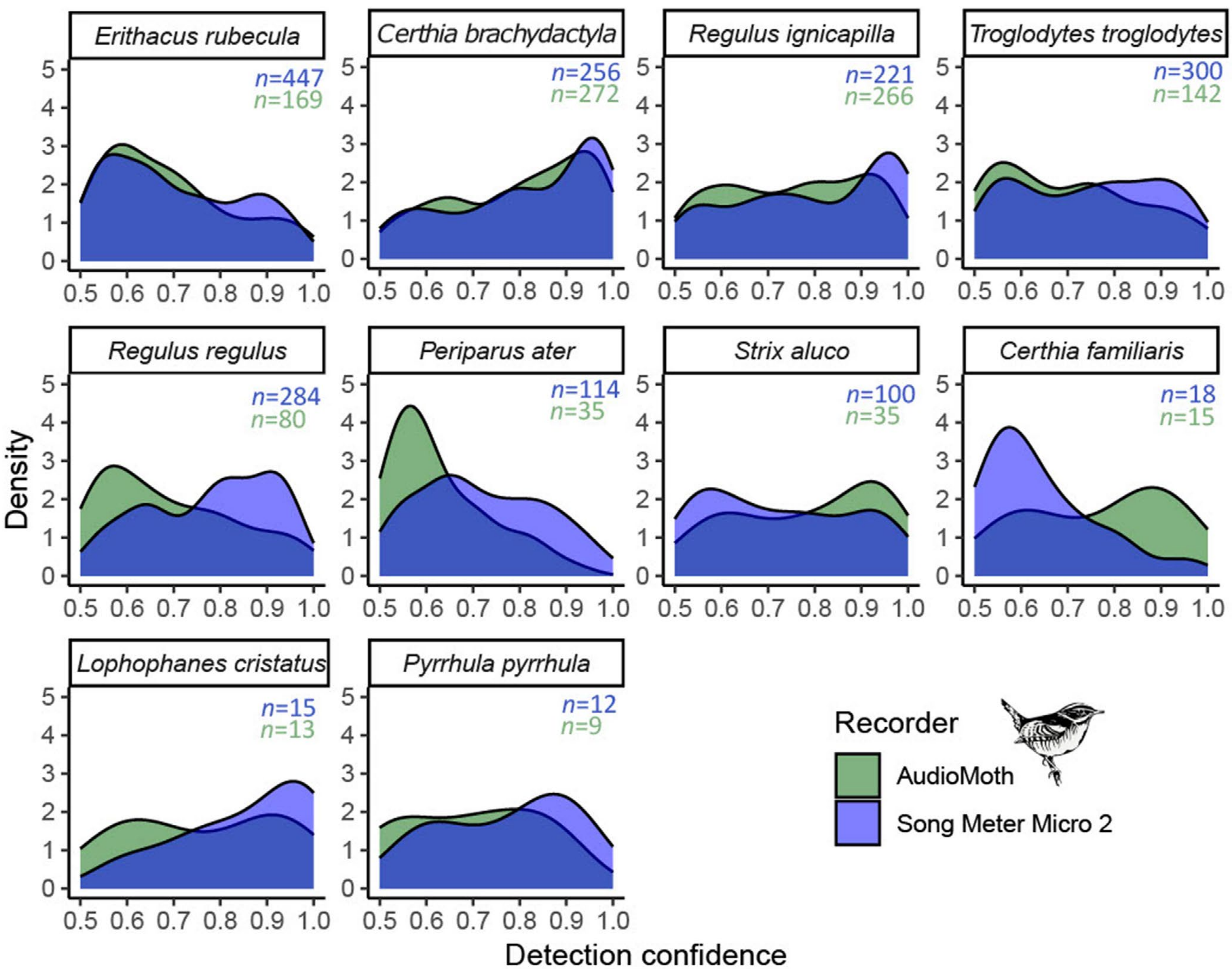

FIGURE 2 Density plots showing the spread of detection confidence scores for the most common species for each recorder. Numbers in each panel refer to the number of detections for that species with each recorder.

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

monitoring, proposed possible solutions, and highlighted the project stage when these biases might be addressed (Tables 2–4). Our case studies demonstrated the ways these biases can cause variation in real-world data and highlighted the challenges researchers and practitioners will face during long-term projects. In the following sections, we provide more detail on practical (Section 4.1) and analytical (Section 4.2) mitigations and summarise the policy implications of the issues covered here (Section 4.3).

4.1 | Practical mitigations

4.1.1 | Equipment

To minimise the need to replace any equipment, such as casing, components or microphones during data collection, it is prudent to purchase spares at the start of the project. Investing in better quality windshields and more robust casing than is obviously needed will counter the possibility of future environmental changes having detrimental effects on recording quality (Table 2). Where equipment change is unavoidable, a period of side-by-side comparison will help quantify differences arising from the transition. Any change would ideally be gradual so that hardware type is not confounded with year, although equipment degradation may make this difficult to achieve (see below).

4.1.2 | Robust monitoring protocols

All hardware, software and analytical pipelines should be documented in the monitoring protocol. When selections are tested before choices are finalised, then the testing methodology and results should also be documented. Collect comprehensive metadata, including all possible confounding variables, so that these can be incorporated into analyses if necessary. When hardware, firmware or software is updated, developers and manufacturers should inform users about the likely consequences of these changes.

4.1.3 | Data storage

Long-term data storage can help mitigate issues caused by software updates or analytical changes because analyses can be rerun on the whole dataset. However, there are substantial financial and environmental costs of data storage: for example, a 20-recorder project recording in wav format for 2h a day at 48kHz would accumulate 96TB of data over 20years, according to the Audiomoth config app (Hill et al., 2019). Long-term storage costs money and data storage facilities are an increasing source of carbon emissions. These consequences can be reduced by retaining only a subset of data, although consideration should be given to the content of the subset: whether to store random samples or a fixed number of positive classifications for example. If storing full audio files is not viable, generative models could offer a means of recovering approximate reconstructions of full audio samples from compressed representations (Gibb et al., 2024). Trade-offs are inevitable, and there is no single option that will fit all long-term monitoring projects (Metcalf et al., 2022).

4.2 | Analytical approaches

4.2.1 | Accounting for bias in models

Where some change in hardware has occurred, the simplest and cheapest means of mitigating bias would be to account for this variation by including 'hardware type' as a factor in the model. For example, when using hierarchical models, hardware type could be included as a random effect, to account for variation between equipment. The success of such an approach would likely be context dependent, with the type of variation, the recording environment, and the timing of the changeover (if the move to new hardware were done gradually over a number of years this may make it easier to statistically account for the effect of the device changeover) determining the extent to which model structure can account for the bias(es). However, if the hardware change coincided with a genuine change in the ecosystem, separating the two might be impossible.

4.2.2 | Calibration testing

Sources of bias caused by many of the issues identified here should be quantified and thus accounted for by periodically measuring deterioration of microphones and effective detection spaces of recorders using a standardised calibration protocol (e.g. Haupert et al., 2023; Yip et al., 2017). The approaches in all these calibration tests are similar and designed to allow the use of correction factors to standardise microphone quality and effective detection space. Yip et al. (2017) played bird vocalisations at known distances and amplitudes and compared detection space of different types of acoustic recorders and a human observer to quantify variation in effective detection distances. Examples using white noise or tones with no comparison to a human observer can be found in Haupert et al. (2023). It is important that calibration tests are designed with consideration for the target signal: if the target species were Goldcrest *Regulus regulus* for example (which vocalises at very high frequencies and is only detectable from very short distances) then a calibration test would need to involve high frequency sounds within the vocal range of this species played from positions near to the recorders. Although speakers will have directionality that differs from the target animal, this directionality will be uniform over time and so calibrating with speakers can effectively measure relative change in microphone performance over time. Where calibration tests reveal a decline (but not complete degradation) in recorder or microphone performance, analyses should take this into account using either an offset for the affected recordings or a random effect for the recorders in the model.

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Large-scale, long-term acoustic projects should build calibration testing into the program. This includes prior to initial recorder deployment and then regularly thereafter, particularly at the end of any field campaign. If microphone design allows, then settings can be changed. If a long-term project has recorders of more than one make and model, ensure the study is designed so that not all units of one type are in a single treatment (if applicable). When developers or manufacturers introduce or modify equipment this should be accompanied by the results of a standardised calibration test.

While the scope of this paper is limited to terrestrial ecoacoustic monitoring, it is worth mentioning that in marine ecoacoustics there is a well-developed literature on underwater calibration testing (e.g. Hayman et al., 2017; Heaney et al., 2020) that incorporates many principles that are applicable to terrestrial monitoring.

4.2.3 | Machine learning

Innovations in deep learning could help mitigate bias arising from changes in hardware used in a long-term project. Simultaneous recordings in the same location with old and new hardware could be modelled to identify residual variation arising from hardware and identify the frequency response and self-noise associated with each hardware type. This would enable a 'mapping between recorders', where the residual variation typical of one hardware type could be added or removed as appropriate. This would account for inter-recorder variation without fundamentally affecting the underlying signal of interest (Gibb et al., 2024). High-resolution reconstruction techniques using generative models may also permit 're-recording' audio captured by previous recorders (where no simultaneous recordings exist), effectively mapping historical samples to the spectral and noise profile of replacement hardware (Duff et al., 2023). This may alleviate some of the challenges of the fast-paced development of acoustic hardware by enabling the reuse of audio from lower quality or partially degraded recording devices. Fieldwork can aid research in this area by collecting simultaneous recordings with multiple different devices and through regular calibration testing of recorders' signal-to-noise ratios (Metcalf et al., 2022).

4.3 | Implications for long-term acoustic monitoring

When designing a long-term acoustic monitoring project, users should identify those biases (Tables 2–4) that are likely to affect their project and build appropriate mitigations into the project design. This should include taking measures to minimise biases likely to arise during recording that could affect the detection space; regular calibration testing to assess that the detection space remains constant over time, and identifying appropriate analytical techniques to account for biases that cause variation in the detection space.

In highlighting the risk of bias in long-term ecoacoustic datasets, we do not intend to dissuade users from attempting long-term monitoring using acoustic methods. Acoustic monitoring is a robust means of data collection and has the potential to be more repeatable and reliable than traditional fieldworker surveys. However, an increased awareness and understanding among hardware manufacturers, researchers, practitioners and policymakers of the challenges, and collaborative efforts to resolve these, will improve the quality of monitoring outputs and benefit the field of ecoacoustics.

AUTHOR CONTRIBUTIONS

Ollie Metcalf conceived the idea and David Jarrett led the writing of the manuscript with support from Tom Bradfer-Lawrence. Ollie Metcalf, Jack Greenhalgh and José Joaquín Lahoz Monfort gathered and analysed the case study data. In addition to the authors named above, Ross Barnett, Jérémy S. P. Froidevaux, Kieran Gibb, Pauline Guinet, Becky Heath, Alison Johnston, Alex Rogers and Stephen G. Willis all contributed critically to reviewing the literature, drafting the paper and gave final approval for publication.

ACKNOWLEDGEMENTS

This paper was developed at the second EcoHack workshop held at the University of Stirling in October 2023. We thank the workshop funders for their support: the Scottish Alliance for Geoscience, Environment and Society, the Company of Biologists (grant number B133), and the UK Acoustics Network (UKAN+, EPSRC grant number EP/V007866/1). D.J. was supported by a Leverhulme Trust Doctoral Studentship grant DS-2017-070 (Interdisciplinary Understanding for a Changing Arctic) and JSPF by a Leverhulme Trust early career fellowship (award reference ECF-2020-571).

CONFLICT OF INTEREST STATEMENT

Jérémy Froidevaux is an Associate Editor of the *Journal of Applied Ecology*, but took no part in the peer review and decision-making processes for this paper. The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

There are no data available with this paper because the case studies are illustrative.

ORCID

David Jarrett https://orcid.org/0000-0001-7600-5447
Ross Barnett https://orcid.org/0009-0000-1392-8404
Tom Bradfer-Lawrence https://orcid.org/0000-0001-6045-4360
Jérémy S. P. Froidevaux https://orcid.org/0000-0001-6850-4879
Kieran Gibb https://orcid.org/0000-0003-1468-547X
Pauline Guinet https://orcid.org/0009-0007-3775-7081
Jack Greenhalgh https://orcid.org/0000-0001-5792-6160
Becky Heath https://orcid.org/0000-0002-9111-6994
Alison Johnston https://orcid.org/0000-0001-8221-013X
Stephen G. Willis https://orcid.org/0000-0002-8656-5808
Oliver Metcalf https://orcid.org/0000-0003-3441-2591

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

Adams, A. M., Jantzen, M. K., Hamilton, R. M., & Fenton, M. B. (2012). Do you hear what I hear? Implications of detector selection for acoustic monitoring of bats. *Methods in Ecology and Evolution*, *3*, 992–998. https://doi.org/10.1111/j.2041-210X.2012.00244.x

Boyd, R. J., Powney, G. D., Burns, F., Danet, A., Duchenne, F., Grainger, M. J., Jarvis, S. G., Martin, G., Nilsen, E. B., Porcher, E., Stewart, G. B., Wilson, O. J., & Pescott, O. L. (2022). ROBITT: A tool for assessing the risk-of-bias in studies of temporal trends in ecology. *Methods in Ecology and Evolution*, *13*(7), 1497–1507.

Boyd, R. J., Powney, G. D., & Pescott, O. L. (2023). We need to talk about nonprobability samples. *Trends in Ecology & Evolution*, *38*(6), 521–531.

Bradfer-Lawrence, T., Desjonqueres, C., Eldridge, A., Johnston, A., & Metcalf, O. (2023). Using acoustic indices in ecology: Guidance on study design, analyses and interpretation. *Methods in Ecology and Evolution*, *14*, 2192–2204. https://doi.org/10.1111/2041-210X.14194

Bradfer-Lawrence, T., Duthie, B., Abrahams, C., Adam, M., Barnett, R. J., Beeston, A., Darby, J., Dell, B., Gardner, N., Gasc, A., Heath, B., Howells, N., Janson, M., Kyoseva, M. V., Luypaert, T., Metcalf, O. C., Nousek-McGregor, A. E., Poznansky, F., Ross, S. R. P. J., … Froidevaux, J. S. P. (2024). The acoustic index user's guide: A practical manual for defining, generating and understanding current and future acoustic indices. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.14357

Bradfer-Lawrence, T., Gardner, N., Bunnefeld, L., Bunnefeld, N., Willis, S. G., & Dent, D. H. (2019). Guidelines for the use of acoustic indices in environmental research. *Methods in Ecology and Evolution*, *10*, 1796–1807. https://doi.org/10.1111/2041-210X.13254

Center for Conservations Bioacoustics. (2023). *Hey, what's new in Raven?* https://updates.ravensoundsoftware.com/updates/raven_pro_1.6_update/whatsnew_ravenpro.html

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

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

Darras, K., Pütz, P., Rembold, K., & Tscharntke, T. (2016). Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biological Conservation*, *201*, 29–37.

Darras, K., Deppe, F., Fabian, Y., Kartono, A. P., Angulo, A., Kolbrek, B., Mulyani, Y. A., & Prawiradilaga, D. M. (2020). High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. *PeerJ*, *8*, e9955.

Dobson, A., Milner-Gulland, E. J., Aebischer, N. J., Beale, C. M., Brozovic, R., Coals, P., Critchlow, R., Dancer, A., Greve, M., Hinsley, A., Ibbett, H., Johnston, A., Kuiper, T., Le Comber, S., Mahood, S. P., Moore, J. F., Nilsen, E. B., Pocock, M. J. O., Quinn, A., … Keane, A. (2020). Making messy data work for conservation. *One Earth*, *22*, 455–465. https://doi.org/10.1016/j.oneear.2020.04.012

Duff, M. A., Simpson, I. J. A., Ehrhardt, M. J., & Campbell, N. D. F. (2023). VAEs with structured image covariance applied to compressed sensing MRI. *Physics in Medicine and Biology*, *68*, 165008. https://doi.org/10.1088/1361-6560/ace49a

Fairbrass, A. J., Rennert, P., Williams, C., Titheridge, H., & Jones, K. E. (2017). Biases of acoustic indices measuring biodiversity in urban areas. *Ecological Indicators*, *83*, 169–177. https://doi.org/10.1016/j.ecolind.2017.07.064

Gibb, K. A., Eldridge, A., Sandom, C. J., & Simpson, I. J. A. (2024). Towards interpretable learned representations for ecoacoustics using variational auto-encoding. *Ecological Informatics*, *80*, 102449. https://doi.org/10.1016/j.ecoinf.2023.102449

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2018). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, *10*(2), 169–185. https://doi.org/10.1111/2041-210X.13101

Hartig, F., Abrego, N., Bush, A., Chase, J. M., Guillera-Arroita, G., Leibold, M. A., Ovaskainen, O., Pellissier, L., Pichler, M., Poggiato, G., Pollock, L., Si-Moussi, S., Thuiller, W., Viana, D. S., Warton, D. I., Zurell, D., & Yu, D. W. (2024). Novel community data in ecology—properties and prospects. *Trends in Ecology & Evolution*, *39*(3), 3227. https://doi.org/10.1016/j.tree.2023.09.017

Haupert, S., Sèbe, F., & Sueur, J. (2023). Physics-based model to predict the acoustic detection distance of terrestrial autonomous recording units over the diel cycle and across seasons: Insights from an Alpine and a Neotropical forest. *Methods in Ecology and Evolution*, *14*, 614–630. https://doi.org/10.1111/2041-210X.14020

Hayman, G., Robinson, S. P., Pangerc, T., Ablitt, J., & Theobald, P. D. (2017). Calibration of marine autonomous acoustic recorders. In *OCEANS 2017 – Aberdeen*, 19–22 June 2017, Aberdeen, UK. https://doi.org/10.1109/OCEANSE.2017.8084773

Heaney, K., Martin, B., Miksis-Olds, J., Ainslie, M., Moore, T., & Warren, J. (2020). ADEON Data Processing Specification, Version 1.0 FINAL. Technical report by Applied Ocean Sciences for Prime Contract No. M16PC00003.

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. https://doi.org/10.1016/j.ohx.2019.e00073

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES Secretariat.

Kahl, S., & Wood, C. (2024). Guidelines for appropriate use of BirdNET scores and other detector outputs. *Journal of Ornithology*, *165*, 777–782. https://doi.org/10.1007/s10336-024-02144-5

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

Lapp, S., Stahlman, N., & Kitzes, J. (2023). A quantitative evaluation of the performance of the low-cost AudioMoth acoustic recording unit. *Sensors*, *23*(11), 5254.

Luna-Naranjo, D., Martinez-Vargas, J. D., Sánchez-Giraldo, C., Daza, J. M., & López, J. D. (2024). Quantifying and mitigating recorder-induced variability in ecological acoustic indices. *Ecological Informatics*, *82*, 102668. https://doi.org/10.1016/j.ecoinf.2024.102668

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M. P., Elston, D. A., Scott, E. M., Smith, R. I., Somerfield, P. J., & Watt, A. D. (2010). Long-term datasets in biodiversity research and monitoring: Assessing change in ecological communities through time. *Trends in Ecology & Evolution*, *25*(10), 574–582. https://doi.org/10.1016/j.tree.2010.06.016

Meng, X. L. (2018). Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics*, *12*(2), 685–726.

Metcalf, O., Abrahams, C., Ashington, B., Baker, E., Bradfer-Lawrence, T., Browning, E., Carruthers-Jones, J., Darby, J., Dick, J., Eldridge, A., Elliott, D., Heath, B., Howden-Leach, P., Johnston, A., Lees, A., Meyer, C., Arana, U. R., & Smyth, S. (2022). Good practice guidelines for long-term ecoacoustic monitoring in the UK. *The UK Acoustics Network*. https://e-space.mmu.ac.uk/id/eprint/631466

Metcalf, O., Baccaro, F., Barlow, J., Berenguer, E., Bradfer-Lawrence, T., Rossi, L. C., do Vale, É. M., & Lees, A. C. (2024). Listening to tropical

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

forest soils. *Ecological Indicators*, *158*, 111566. https://doi.org/10.1016/j.ecolind.2024.111566

Miquel, J., Latorre, L., & Chamaillé-Jammes, S. (2022). Addressing power issues in biologging: An audio/inertial recorder case study. *Sensors*, *22*, 8196. https://doi.org/10.3390/s22218196

Obrist, M. K., Pavan, G., Sueur, J., Riede, K., Llusia, D., & Márquez, R. (2010). Bioacoustics approaches in biodiversity inventories. *Abc Taxa*, *8*, 68–99.

Open Acoustic Devices. (2024). Application notes. https://www.openacousticdevices.info/application-notes

Perea, S., & Tena, E. (2020). Different bat detectors and processing software... same results? *Journal of Bat Research & Conservation*, *13*, 4–8. https://doi.org/10.14709/BarbJ.13.1.2020.01

Pereira, H. M., & Cooper, H. D. (2006). Towards the global monitoring of biodiversity change. *Trends in Ecology & Evolution*, *21*(3), 123–129. https://doi.org/10.1016/j.tree.2005.10.015

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.

R Core Team. (2023). *R A language and environment for statistical computing*. R Foundation for Statistical Computing.

Roe, P., Eichinski, P., Fuller, R. A., McDonald, P. G., Schwarzkopf, L., Towsey, M., Truskinger, A., Tucker, D., & Watson, D. M. (2021). The Australian acoustic observatory. *Methods in Ecology and Evolution*, *12*, 1802–1808. https://doi.org/10.1111/2041-210X.13660

Sánchez-Giraldo, C., Bedoya, C. L., Morán-Vásquez, R. A., Isaza, C. V., & Daza, J. M. (2020). Ecoacoustics in the rain: Understanding acoustic indices under the most common geophonic source in tropical rainforests. *Remote Sensing in Ecology and Conservation*, *6*, 248–261. https://doi.org/10.1002/rse2.162

Schmeller, D. S., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., Déri, E., Budrys, E., Kull, T., Tali, K., Bauch, B., Settele, J., Van Swaay, C., Kobler, A., Babij, V., Papastergiadou, E., & Henle, K. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, *23*, 307–316. https://doi.org/10.1111/j.1523-1739.2008.01125.x

Sethi, S. S., Fossøy, F., Cretois, B., & Rosten, C. M. (2021). *Management relevant applications of acoustic monitoring for Norwegian nature—The sound of Norway. NINA report 2064*. Norwegian Institute for Nature Research.

Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave: A free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*, 213–226. https://doi.org/10.1080/09524622.2008.9753600

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. https://doi.org/10.1093/biosci/biy147

Thomas, A., 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 - Austral Ornithology*, *120*, 239–248. https://doi.org/10.1080/01584197.2020.1787848

Turgeon, P., Van Wilgenburg, S. L., & Drake, K. L. (2017). Microphone variability and degradation: Implications for monitoring programs employing autonomous recording units. *Avian Conservation and Ecology*, *12*(9). https://doi.org/10.5751/ACE-00958-120109

Villanueva-Riviera, L. J., & Pijanowski, B. C. (2018). Soundecology: Soundscape ecology. https://cran.r-project.org/web/packages/soundecology/index.html

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New.

Wildlife Acoustics. (2024). Release notes. https://www.wildlifeacoustics.com/release-notes/kaleidoscope-pro

Yip, D. A., Leston, L., Bayne, E. M., Sólymos, P., & Grover, A. (2017). Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conservation and Ecology*, *12*(1), 11.

Zhang, C., Zhang, Y., Zheng, X., Gao, X., & Hao, Z. (2024). Influence of recording devices and environmental noise on acoustic index scores: Implications for bird sound-based assessments. *Ecological Indicators*, *159*, 111759. https://doi.org/10.1016/j.ecolind.2024.111759

How to cite this article: Jarrett, D., Barnett, R., Bradfer-Lawrence, T., Froidevaux, J. S. P., Gibb, K., Guinet, P., Greenhalgh, J., Heath, B., Johnston, A., Monfort, J. J. L., Rogers, A., Willis, S. G., & Metcalf, O. (2025). Mitigating bias in long-term terrestrial ecoacoustic studies. *Journal of Applied Ecology*, *00*, 1–12. https://doi.org/10.1111/1365-2664.70000

13652664, 0, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.70000 by Durham University - University, Wiley Online Library on [14/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License