Imperial College London

Dyson School of Design Engineering

On the spatial and resilience-related features of natural soundscapes

Neel Patrick Le Penru

October 2025

Revised May 2026

Supervised by

Prof. Lorenzo Picinali (Dyson School of Design Engineering)

Prof. Robert M. Ewers (Department of Life Sciences)

Dr. Sarab S. Sethi (Department of Life Sciences)

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

Ecosystems worldwide are under unprecedented, intensifying pressure from human activity and climate change, with profound implications for nature and society. To understand and mitigate the ecological impacts of these global crises, we need accurate, scalable, responsive tools to monitor ecosystem health and biodiversity. Ecoacoustic monitoring is a powerful and versatile method to this end, owing to the information-rich nature of soundscapes – comprising sounds from animals, humans, other organisms and environmental noise – and cheaper, more capable autonomous recording hardware and analysis software. In this thesis, I develop and apply novel technologies for acoustically monitoring some of the world's most biodiverse – and vulnerable – ecosystems, with a focus on two twin strands: leveraging the spatial distribution of sounds for enhanced biodiversity assessment with multichannel recording and analysis, and exploring whether acoustic monitoring can detect early warning signals of declining ecosystem resilience (capacity to recover from perturbation) towards tipping points – nonlinear shifts to alternative, often undesirable states triggered by small changes or perturbations. I first investigate ambisonic soundscape reconstruction as a lifelike laboratory test environment for spatial ecoacoustic technologies. I evaluate the effectiveness of this method by using it to test the Multichannel Acoustic Autonomous Recording Unit (MAARU), a custom, six-microphone recorder. I then refine MAARU's design, co-supervise its deployment in Brazil, and deploy it in the UK, USA and Malaysia to investigate its potential to improve avian call classification and distinguish neighbouring habitats via beamforming, which amplifies sounds from target directions. Finally, I explore satellite and acoustic data from Malaysia and Brazil for early warning signals of rainforest to savannah tipping points. I close by considering how these technologies advance our ability to understand and intervene in rapid ecological change, and relate the overall work to the vital need for ecosystem, individual and societal resilience – Earth's future depends upon it.

Statement of Originality

I hereby declare that work in this thesis is my own and that the appearance or discussion of any work from others is appropriately referenced in accordance with standard practices. As is common in this discipline, some parts of the research were conducted collaboratively (for instance, advice from supervisors and collaborations with fellow students and colleagues) and are duly acknowledged.

NB. The Brazil data used in Chapter 5 were shared by Oliver Metcalf, who gathered it with colleagues from the Sustainable Amazon Network. Other shared data are acknowledged as appropriate in later sections.

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

I am hugely grateful to my supervisors, Lorenzo Picinali, Rob Ewers and Sarab Sethi. Thank you for your continual support, wisdom and patience – it has been a privilege to work with you.

I have been fortunate to collaborate with some brilliant post-doctoral researchers who supported me in my work: Becky Heath, Oliver Metcalf, Tom Bury and Jamie Dunning. Thank you for your contributions and kind-heartedness. I have also been very lucky to work with several talented students and recent graduates to whom I extend my thanks: Shujie Chang, Giulia Gieve, James Skinner, James Garside and Alessia Borrelli.

Some of the best and most rewarding moments of this PhD took place abroad, and I am indebted to all those who made me feel so welcome during these visits. From Cornell, I thank Holger Klinck, Irina Tolkova, Wendy Erb, Jordan Aceto, Larissa Sugai, Ashik Rahaman, Ana Verahrami, Isla Keesje, Patrick Chaopricha, Connor Wood, Ben Gottesman, Ben Mirin, Isabella Marshall, Lisa Yang, Laurel Symes, Sharon Martinson and many others. I hope to see you all back in Ithaca someday. From Malaysia, I owe thanks to my local collaborator, Henry Bernard, as well as a brilliant logistical and field support team from SEARRP: Ahmad Jelling, Jamiluddin Jami, Adrian Karulos, Johnny Larenus, Yehezkiel Jahuri, and Louis Makhluf. I similarly hope my field trip will not have been my last visit to the Danum Valley. I also thank the Sabah Biodiversity Center for an access licence and the Danum Valley and Maliau Basin Management Committees for providing fieldwork permits.

I must thank the members past and present of the Audio Experience Design and Ewers groups for always providing support when needed and making the day-to-day so inspiring. Thanks to: Sui Peng Heon, Rapolas Daugintis, David Orme, Kat Poole, Oliver Turvey, Tijmen de Lorm, Jacob Cook, Vivienne Groner, Julie Meyer, Hollie Folkard-Tapp, Thibault Vicente, Johan Pauwels, Aidan Hogg, and a number of others.

Much of this work wouldn't have been possible without some very dedicated administrative teams keeping things running in the background. Here too there are too many mention, but I must thank Christiane Morgan, Anne Houston, Alex Gibbs, Olivia Daniel, Nicholas Moult, Hala Khateeb, Duncan Walsh, and Catalina Estrada Montes amongst others. In addition, I must thank the ACE Workshop team, including Ingrid Logan, Chloe Allen-Greeves, Tito Nwofor, Gordon Addy and Claudia Morgan, as well as the IDEAS lab student mentors and Paul Beasley at Silwood Park for all helping my prototype recorders come to life.

I am also grateful to my PhD examiners, Dan Stowell and Thrishantha Nanayakkara, whose constructive comments helped strengthen many aspects of this thesis.

This work was funded by the UK Natural Environment Research Council via the Science and Solutions for a Changing Planet DTP. Additional travel grants for fieldwork and conferences were provided by the UK Acoustics Network Plus, the UK Turing Scheme (via the Imperial Global Fellows Fund), the Imperial College Trust and the Old Centralians' Trust.

Last but certainly not least, I am so, so lucky to have such wonderful and caring friends, bandmates and family. Your support and good humour have kept me going. I especially thank my mum, Manjeet, my dad, Patrick, my brother, Mathieu, my grandmothers Nanima and Mamie Ginette, and Annette. Thank you all – your unending love and support mean the world to me. This thesis is for you.

Publications and Contributions

Prospects for acoustically monitoring ecosystem tipping points [Preprint]
N. P. Le Penru, T. M. Bury, S. S. Sethi, R. M. Ewers, L. Picinali
arχiv. 2025; https://doi.org/10.48550/arXiv.2509.02201
Contribution: Led the review of literature and writing of the manuscript.
Appearance in thesis: Chapter 2 (with minor edits)
Acknowledgements: T. M. Bury wrote the derivation of variance and autocorrelation as early warning signals (Section A.1) and helped revise the content in Table A.1. All co-authors contributed to revising the text.

Towards using virtual acoustics for evaluating spatial ecoacoustic monitoring technologies [Journal Article]
N. P. Le Penru, B. E. Heath, J. Dunning, L. Picinali, R. M. Ewers, S. S. Sethi
Methods in Ecology and Evolution. 2025; https://doi.org/10.1111/2041-210X.14405
Contribution: Led the design of the virtual sound environment, fieldwork, data collection, analysis and writing of the manuscript, except for tasks acknowledged below.
Appearance in thesis: Chapter 3 (with minor edits)
Acknowledgements: J. Dunning performed the manual labelling of avian calls in recordings, and B. E. Heath ran the recordings through HARKBird. All co-authors contributed to revising the text. A. Borrelli (undergraduate research placement student) contributed to the fieldwork and data collection.

Development and validation of a test platform for spatial ecoacoustic technologies [Conference Paper]
N. P. Le Penru, A. Borrelli, B. E. Heath, R. M. Ewers, S. S. Sethi, L. Picinali
In: *Proceedings of Forum Acusticum 2023, 10th Convention of the European Acoustics Association*, eds. A. Astolfi, F. Asdrubali, L. Shtrepi. Turin, Italy.
2023; https://doi.org/10.61782/fa.2023.0339
Contribution: Precursor to 'Towards using virtual acoustics for evaluating spatial ecoacoustic monitoring technologies'; see above.
Appearance in thesis: Chapter 3 (in edited form)
Acknowledgements: As above.

Insights from deploying Multichannel Acoustic Autonomous Recording Units in Borneo and the UK [Conference Paper]
N. P. Le Penru, B. E. Heath, R. M. Ewers, S. S. Sethi, L. Picinali
In: *Proceedings of Forum Acusticum Euronoise 2025, 11th Convention of the European Acoustics Association*, eds. D. de la Prida, J. Ramis, M. Machimbarrena. Málaga, Spain. 2025; https://www.doi.org/10.61782/fa.2025.0480
Contribution: Led the redesign of recorder weatherproofing, construction of most recorder units, fieldwork and writing of the manuscript.
Appearance in thesis: Chapter 4, Section 4.2 (in edited form)
Acknowledgements: Recorders were built with assistance from fellow PhD students in the Audio Experience Design group and technicians from the Imperial ACE Workshop and IDEAS Lab. P. T. Chaopricha at the Cornell Lab of Ornithology suggested using metal mesh on the exterior of the casing to prevent insect ingress via microphone vents. UK fieldwork was supported by S. Chang and Silwood Park technician P. Beasley. Borneo fieldwork was supported by team members and research assistants at the South East Asia Rainforest Research Partnership (SEARRP; see general thesis Acknowledgements), an access licence from the Sabah Biodiversity Centre and a fieldwork permit from the Danum Valley Management Committee. All co-authors contributed to revising the text.

Spatial ecosystem monitoring with a Multichannel Acoustic Autonomous Recording Unit (MAARU) [Journal Article]
B. E. Heath, R. Suzuki, **N. P. Le Penru**, J. Skinner, C. D. L. Orme, R. M. Ewers, S. S. Sethi, L. Picinali
Methods in Ecology and Evolution. 2025; https://doi.org/10.1111/2041-210X.14390
Contribution: Led some of the recorder localisation tests and revised the manuscript. Co-supervised (with L. Picinali) and collaborated with J. Skinner (then Master's student) on the following: 3D-printed recorder weatherproofing, recorder software bug fixes, UK fieldwork, conceptualisation and interpretation of results from arbitrary-direction beamforming study.
Appearance in thesis: Chapter 4, Section 4.4 (in edited form)
Acknowledgements: In addition to the collaborative tasks noted above, J. Skinner led the fieldwork in Brazil and analysis of beamforming in arbitrary static directions, and facilitated the introduction to AMAgroup. AMAgroup provided financial and logistical support for the Brazil fieldwork.

Accelerating loss of resilience in Bornean rainforests [Preprint]
S. Chang, **N. P. Le Penru**, R. M. Ewers
bioRχiv. 2025; https://doi.org/10.1101/2025.07.21.665904
Contribution: Conceptualised and co-supervised the study with R. M. Ewers. Co-

designed and revised analyses and figures, and co-wrote manuscript with S. Chang (then Master's student).

Appearance in thesis: Chapter 5 (in edited form)

Acknowledgements: S. Chang implemented all satellite data analyses and associated figure plotting in R, and reviewed additional literature. All co-authors contributed to revising the text.

Contents

Glossary **13**

List of Common Abbreviations **20**

List of Figures **22**

List of Tables **25**

1 Introduction: Into a world of sound **26**
1.1 Key Concepts . . . 28
1.1.1 Ecoacoustics: The Study of Soundscapes . . . 28
1.1.2 Spatial Audio and Multichannel Recording . . . 29
1.1.3 Feedbacks Govern the World . . . 31
1.2 Research Objectives . . . 32
1.3 Thesis Outline . . . 33

2 Literature Review: Prospects for acoustically monitoring ecosystem tipping points **36**
2.1 Introduction . . . 37
2.1.1 Box 1: Basic Theory of Bifurcation Tipping Points . . . 40
2.2 Step-by-Step Considerations for Acoustically Identifying Tipping Points . . . 44
2.2.1 Step 1: Will it Tip? Identifying Tipping Potential . . . 44
2.2.2 Step 2: Gathering Data at Sufficient Scale and Resolution . . . 45
2.2.3 Step 3: Extracting Relevant Ecological Metrics from Acoustic Data 46
2.2.4 Step 4: Data Pre-Processing . . . 50
2.2.5 Step 5: Searching for Early Warning Signals . . . 52
2.2.6 Box 2: Applied Example . . . 55
2.3 Opportunities and Challenges for Acoustically Monitoring Tipping Point Early Warning Signals . . . 57
2.3.1 Acoustics May Reveal Early Warning Signals at Various Ecosystem Levels . . . 57
2.3.2 Acoustic Data are Limited but Growing . . . 58
2.3.3 Uncertainties in Tipping Point Theory and Early Warning Signals . 60
2.4 Conclusion . . . 61

3 Towards using virtual acoustics for evaluating spatial ecoacoustic technologies **62**
3.1 Introduction . . . 63
3.2 Materials and Methods . . . 68
3.2.1 Platform Description . . . 68

3.2.2 Recording Devices, Locations and Schedule . . . 70
3.2.3 Soundscape Reproduction Efficacy and Impact of Device Orientation 71
3.2.4 BirdNET and HARKBird Performance on Additional Avian Calls . 73
3.3 Results . . . 75
3.3.1 Soundscape Reproduction Efficacy and Impact of Device Orientation 75
3.3.2 BirdNET and HARKBird Performance on Additional Avian Calls . 82
3.4 Discussion . . . 83
3.4.1 Soundscape Reproduction Efficacy and Impact of Device Orientation 84
3.4.2 BirdNET and HARKBird Performance on Additional Avian Calls . 86
3.5 Guidelines and Recommendations . . . 87

4 Spatially segmenting soundscapes with a refined multichannel autonomous recorder for enhanced biodiversity assessment 89
4.1 Introduction . . . 90
4.2 Sub-Project 1: MAARU weatherproofing redesign and field robustness tests 95
4.2.1 Methods . . . 95
4.2.2 Results . . . 98
4.2.3 Discussion . . . 98
4.3 Sub-Project 2: Theoretical Array Directivity . . . 99
4.3.1 Background on Array Geometry and Directivity . . . 99
4.3.2 Methods . . . 101
4.3.3 Results . . . 101
4.3.4 Discussion . . . 103
4.4 Sub-Project 3: Initial Beamforming Validation . . . 104
4.4.1 Methods . . . 104
4.4.2 Results . . . 105
4.4.3 Discussion . . . 106
4.5 Sub-Project 4: Beamforming in Directions of Ecological Interest . . . 107
4.5.1 Methods . . . 107
4.5.2 Results . . . 110
4.5.3 Discussion . . . 118
4.6 Outlook and Conclusions . . . 121

5 Multimodal assessment of tropical rainforest resilience loss 123
5.1 Introduction . . . 124
5.2 Materials and Methods . . . 127
5.2.1 Satellite Data Analysis . . . 127
5.2.2 Acoustic Data Analysis . . . 131
5.3 Results . . . 136
5.3.1 Satellite-Sensed Vegetation Resilience . . . 136
5.3.2 Correlates of Vegetation Resilience . . . 141
5.3.3 Nascent Early Warning Signals in Acoustic Data . . . 142
5.4 Discussion . . . 146
5.4.1 Satellite-Sensed Vegetation Resilience . . . 146
5.4.2 Correlates of Vegetation Resilience . . . 147
5.4.3 Nascent Early Warning Signals in Acoustic Data . . . 149
5.5 Outlook and Conclusions . . . 152

6 Conclusion: Raising alarms and sounding change 154

6.1 Synthesis . . . 154
6.2 Evaluation of Research Objectives . . . 155
6.3 Future Directions . . . 156
6.3.1 Where Next for MAARU? . . . 156
6.3.2 Realising Multimodal Tipping Point Monitoring with (Spatial) Ecoacoustics . . . 157
6.3.3 Positive Tipping Points . . . 159

Bibliography **161**

A Supplementary Material for Chapter 2 **204**
A.1 Derivation of variance and autocorrelation as tipping point EWS . . . 210
A.1.1 Continuous-time process . . . 210
A.1.2 Discrete-time process . . . 212

B Supplementary Material for Chapter 3 **215**

C Supplementary Material for Chapter 4 **220**

D Supplementary Material for Chapter 5 **225**

Glossary

Acoustic Indices Numerical summaries of acoustic data designed to characterise the biophony, anthropophony and geophony in soundscapes based on various spectral and temporal features of recordings. Acoustic indices are generally univariate and often based on the distribution of energy within recordings. They have provided a rapid means of analysing soundscapes from diverse ecosystems in ecoacoustic studies, but their correlations to biodiversity and ecological patterns are often inconsistent and study-specific. As such, they should be used and calibrated carefully.

Aliasing In acoustics, (temporal) aliasing is the introduction of 'grating' (i.e., unwanted) frequency lobes in a signal's spectrum as a result of temporally sub-sampling it into the digital domain (specifically, sampling it below twice its Nyquist frequency), such that it contains erroneous sounds and/or sounds distorted.

Ambisonics A method for the flexible and efficient recording and reproduction of 3D spatial sound fields (i.e., spatial audio recording and playback). Ambisonic recording involves encoding multi-microphone recordings to a format in which sound fields are decomposed into spatial components using a set of mathematical functions known as spherical harmonics. Ambisonic playback involves decoding from this format using loudspeaker arrays to recreate the spatial sound field in a perceptually lifelike manner.

Anthropophony The sounds produced by human activity (e.g., speech, machinery operation, vehicle use).

Autonomous Recording Unit A sensor-equipped recorder designed to independently gather data over extended periods in the field. This thesis focusses primarily on autonomous recording units with microphones for passive acoustic monitoring, though units with other sensors (e.g., cameras) also exist. Autonomous recording units facilitate large-scale passive acoustic monitoring and can either be semi-autonomous, requiring occasional manual intervention (e.g., to replace batteries and retrieve data), or fully-autonomous with self-sufficient powering and data transmission (e.g., solar-powering and 4G network connectivity).

Basin of Attraction A set of system state values about a stable equilibrium, and associated parameter values (i.e., conditions), from which the system returns to that equilibrium. The basin of attraction is often illustrated using the potential landscape of a system and the related 'ball and cup' model, in which the basin of attraction is the well or 'cup' around a stable equilibrium and the system state is a ball moving along potential landscape's surface. In the basin of attraction, the ball usually rolls back to the bottom of the cup (the stable equilibrium point) unless it is pushed out

of it or the cup disappears due to a bifurcation. See Section 2.1.1, Fig 2.1 and Table A.1 for further details.

Beamforming A class of spatial filtering techniques that can be applied to multichannel recordings to amplify sounds from target directions and reduce background noise. These techniques often take the weighted sum of each channel's recording so that correlated (i.e., predominantly in-phase) signals like bird calls from a unique direction constructively superpose and decorrelated diffuse noise is attenuated by destructive interference. Beamforming can therefore be used to 'steer' multichannel recorders to focus on particular directions, commonly termed 'look directions'.

Bifurcation A qualitative or topological change to the dynamics of a system in response to smooth, gradual change of its control parameter(s) (i.e., conditions) passing a critical threshold known as the 'bifurcation point'. Such qualitative changes can include the loss of stability at a local equilibrium (in the case of 'local bifurcations'), and the creation or annihilation of fixed points (stable states). Bifurcations can be grouped into various prototypical 'normal forms', several of which describe archetypal transitions between stable states that underpin tipping points in many natural systems. 'Fold bifurcations' in particular are associated with numerous tipping points for abrupt transitions in the climate and ecosystems (e.g., lakes switching from clear to turbid, forests transitioning to savannah, and the collapse of ice sheets, ocean circulation, overharvested populations, and more). They describe a stable state's gradual loss of stability as conditions (e.g., temperature, precipitation, harvesting rates) change, leading to an eventual collision with an unstable state that results in a sudden, discontinuous shift to a contrasting stable state. See Fig 2.1, Section 2.1.1 and Table A.1 for further details.

Bioacoustics The study of sonic animal communication and/or sounds produced by specific species or organisms. The precise definition and use of the term can vary in different contexts.

Biophony The sounds produced by living organisms and their interactions with their environment (e.g., bird calls, cricket stridulations, the buzz of bees).

BirdNET A convolutional neural network-based machine learning classifier that (primarily) predicts bird species based from recordings of their vocalisations. BirdNET is freely-available, can detect over 6,000 species and was trained on substantial amounts of data including 100,000s recordings from large public sound repositories of bird recordings, Xeno-Canto and the Macaulay Library. It is also capable of classifying certain other species, such as anurans, and was trained to detect 11 'non-event' classes (e.g., human voices, dogs, engines, and guns). It is one of the most popular models for species call classification and widely used in passive acoustic monitoring studies across the globe.

Breakpoint In this thesis, 'breakpoint' refers to the time at which the linear trend in a time series changes. This is determined by applying linear regressions to different segments of the time series using Bayesian inference. The slopes of linear regressions fit to the data for periods before and after the breakpoint summarise the magnitude of change in each segment.

Centre Manifold Theorem Theorem that states that when large, complex dynamical systems approach a local bifurcation, their dynamics simplify to one of various lower-dimensional bifurcation normal forms (e.g., fold bifurcation, transcritical bifurcation).

Clock Drift Inaccuracies that gradually accumulate in the timing clock used to control the temporal sampling of signals in recorders. Clock drift leads to delays between recorders, which can cause signal source localisation or direction of arrival estimation based on differences in time of arrival at multiple independent recorders to fail. It can be avoided by synchronising recorders (e.g., with GPS) or having multiple sensors share the same clock, for instance in a multichannel recorder with an integrated microphone array.

Critical Slowing Down The slowing of a system's return time to its stable equilibrium as changing conditions erode its stability (dominant eigenvalue, $\lambda \to 0$) towards a tipping point for a bifurcation. Such slowing down results in the system state exhibiting longer and larger fluctuations about its equilibrium due to noise or perturbations. This pattern can be detected in various ways, providing a basis for various early warning signals of bifurcation tipping points. For further explanation, see Section 2.1.1 and Table A.1.

Critical Transition General term for abrupt transitions to a contrasting state in complex dynamical systems, such as the climate, ecosystems, financial markets and social systems, amongst others. While 'tipping point' denotes the point at which such transitions can be induced by small change, 'critical transition' refers to the transitions themselves.

Diffuse Sounds Diffuse sounds are those whose energy is distributed near-uniformly in a sound field, such that at any point in space they appear to arrive from all directions. Examples include the stridulation of insects distributed approximately evenly throughout an environment, or sounds that are scattered around a space through numerous reflections on different surfaces.

Direction of Arrival The direction from which a signal originates relative to a sensor. Integrated microphone arrays can be used to estimate the direction of arrival of acoustic signals.

Directionality The degree to which a signal transmits along a precise direction and/or appears to originate from a unique location. In acoustics, lower frequencies are less directional than higher frequencies, and diffuse sounds are the opposite of directional sounds.

Distributed Microphone Array A large array of distantly-spaced (typically on the order of tens of centimetres to tens of metres) microphones. These may be connected to the same recorder via long cables, or record onto independent recorders (though this poses the risk of clock drift).

Early Warning Signal Indicator of an impending tipping point or critical transition in a system. Common temporal early warning signals include increases in certain statistical properties of time series data of a system's state indicative of critical slowing down, namely rising autocorrelation and variance (see Table A.1 and Section A.1 for further details).

Ecoacoustics The study of natural soundscapes to monitor biodiversity and ecosystem health. In contrast to bioacoustics, ecoacoustics typically looks at an environment's entire soundscape (or a wider part of it), in order to study questions that relate to ecosystem function and can address any level of ecological organisation, including populations, communities and entire landscapes. It thus encompasses the study of 'soundscape ecology' which examines the relationship between soundscapes and landscapes.

Ecological Resilience Ecological resilience corresponds to the amount of disturbance a ecosystem can endure without changing its function or state. Although various factors may contribute to ecological resilience, it depends fundamentally on the shape and size of the basin of attraction about an ecosystem's stable state, which are mathematically related to engineering resilience and the system's dominant eigenvalue. See Section 2.1.1, Fig 2.1 and Table A.1 for further details.

Engineering Resilience Engineering resilience corresponds to how fast a system recovers from perturbation (with faster recovery indicating higher resilience). As such, it depends on the mathematical definition of stability for dynamical systems. See Section 2.1.1 and Table A.1 for further details.

Geophony The sounds of the geophysical processes in an environment (e.g., wind, rain, thunder).

HARKBird A freely-available software tool for estimating the direction of arrival bird calls in multichannel recordings made with integrated microphone arrays. It is based on the robot audition software, HARK (Honda research institute Japan Audition for Robots with Kyoto University) and uses the MUSIC (MUltiple SIgnal Classification) algorithm for direction of arrival estimation.

Hysteresis A phenomenon characterised by a delay in the response of a system's state (or some property) to changing conditions. Hysteresis is a defining characteristic of fold bifurcations, where, in practical terms, reverting a system to its prior state following a transition requires changing conditions much further back beyond the threshold that brought about the current, new state.

Integrated Microphone Array A small array of closely-spaced (typically on the order of centimetres) microphones connected to a single recorder.

Kendall's τ A non-parametric rank correlation coefficient that examines trends (specifically, tendency) between two variables. For time series data (where one variable is time), a Kendall's Tau value of 1 indicates the data are always increasing over time, -1 indicates the data are always decreasing, and 0 indicates the lack of an overall trend. Kendall's τ is widely used in tipping points research to assess whether statistical early warning signals are increasing or decreasing overall.

Masking A phenomenon whereby a target sound is obscured by overlapping sound(s) occurring at the same time. Often, the target and overlapping sounds originate from different directions, so one approach to tackling masking is to directionally filter audio using multichannel recorders and beamforming.

Model Embeddings The numerical outputs of an internal layer or representation of data in a trained machine learning model or larger and more complex deep learning model. In ecoacoustics, model embeddings are often extracted from acoustic classifiers (e.g., BirdNET) and typically correspond to sets of hundreds or thousands of numbers that describe a recorded soundscape and may serve as powerful proxies for ecosystem health and biodiversity. In this sense, model embeddings may be considered as multivariate 'machine-learned' acoustic indices, as opposed to the univariate 'human-derived' acoustic indices that have traditionally been used to characterise soundscapes and biodiversity.

Multichannel A multichannel recorder (or transmitter) contains multiple sensors (or antennae). Multichannel recordings from a sensor array contain multiple channels of data (one for each sensor). In this thesis, multichannel (acoustic) recorders are those with integrated microphone arrays.

MUSIC Algorithm The MUltiple SIgnal Classification (MUSIC) algorithm is a widespread method for estimating signal directions of arrival in multichannel sensor array data. It works by taking an eigenvalue decomposition of the multichannel signal covariance matrix and separating its eigenvectors into two subspaces: one for eigenvectors assumed to correspond to signal energy (signal subspace) and one for eigenvectors assumed to correspond to noise (noise subspace). Array steering vectors (which describe how signals from given directions are received across the sensor array) for signal directions of arrival should have a minimal projection onto the noise subspace. The algorithm therefore calculates the reciprocal of the steering vector's squared projection on the noise subspace over the full range of possible azimuth and elevation angles. Peaks in this function are then used as direction of arrival (azimuth and elevation) estimates for each target signal. Key assumptions of the MUSIC algorithm are that there are fewer target signals than there are receivers (e.g., microphones) in the sensor array, and that target signals are uncorrelated.

Passive Acoustic Monitoring The deployment of acoustic autonomous recording units in terrestrial or aquatic ecosystems over extended periods of time (and increasingly, over large spatial scales), in order to record natural soundscapes to monitor biodiversity and ecosystem health.

Pitch Angle An object's angle of rotation about its horizontal lateral (a.k.a. transverse) axis. In aircraft, the pitch angle determines whether a plane's nose faces up or down.

Potential Landscape A surface that represents the dynamics of a system in relation to the system state, with peaks at unstable equilibria and valleys at stable equilibria — the latter correspond to 'basins of attraction'. In certain systems (e.g., some mechanical systems), the height of the surface corresponds to potential energy. Although the notion of 'potential' is more abstract for ecological systems, potential landscapes are often used to illustrate and quantify aspects of stability, resilience and alternative stable states in natural systems. See Section 2.1.1, Fig 2.1 and Table A.1 for further details.

Regime Shift All-encompassing term for shifts between alternative states in dynamical systems. Although the term 'regime shift' is often used interchangeably with 'critical transition' and 'tipping point', it specifically seeks to capture the idea that

real-world systems generally do not stay still or fixed at their stable states. Rather, stable states in practice correspond to 'dynamic regimes', where local dynamics may involve continual fluctuations about an equilibrium due to noise or perturbations, or, more broadly, various types of trajectories about an attractor (e.g., cyclic or chaotic dynamics). Regime shifts thus denote transitions between alternative dynamic regimes in systems.

Soniferous Sound-producing.

Sound Truth The practice of listening to acoustic soundscape recordings to understand their content and validate the outputs of ecoacoustic analyses. 'Sound truthing' is analogous to ground truthing, but only involves inspecting recordings rather than collecting real-world validating data in the field (e.g., via manual field surveys or trusted environmental sensing systems).

Soundscape The amalgamation of all sounds in an environment, comprising biophony, geophony and anthropophony.

Spatial Aliasing The introduction of 'grating' (i.e., unwanted) lobes in the directivity power pattern of a sensor array for different frequencies, as a result of spatially subsampling signals of those frequencies. The frequency threshold above which spatial aliasing occurs is dictated by the spatial sampling theorem, a spatial analogue of the Nyquist-Shannon theorem (Section 4.3.1). It states that the maximum distance between sensors must be half the wavelength of the threshold frequency (so that there are at least two sensors per wavelength of the maximum un-aliased frequency). For signals above the frequency threshold, spatial aliasing can result in those from non-target directions being detected when beamforming or can lead to errors in their direction of arrival estimates. Spatial aliasing can also lead to distortions in the source position/direction of signals in spatial sound fields reproduced via virtual sound environments.

Stability In ecology, stability is often viewed as a multifaceted property that relates to different features of how populations, communities or ecosystems resist and recover from perturbation or disturbance. This builds upon the definition of stability for dynamical systems, which considers a system's equilibrium state to be stable if trajectories after perturbation return to it. Mathematically, the dynamical systems theory definition of stability depends on the dominant eigenvalue of the system's Jacobian matrix (λ). If the real part of the dominant eigenvalue is negative, the system's equilibrium is stable. The dominant eigenvalue also determines the rate at which the system will return to equilibrium (if stable) or move away from it (if unstable). For further explanation, see Section 2.1.1 and Table A.1.

Tipping Point The point at which a small change to the state of a system or an increment in conditions can lead to a sudden, non-linear transition amplified by positive feedback, such as the abrupt shift from one stable state to another. Tipping points have been closely associated with 'fold bifurcations', but they can be brought about by other forms of bifurcation, internal noise shifting a system out of its basin of attraction, or rapid forcing (i.e., change) of the system conditions.

Vegetation Optical Depth A microwave-based satellite sensor product that measures the water content of vegetation from space and serves as a proxy for vegetation

biomass. Different microwave wavelengths (denoted as different VOD 'bands', e.g., the Ku-band) tend to reflect from water in different parts of vegetated areas and may thus be more sensitive to water content in e.g., soil moisture or forest canopy cover.

Virtual Sound Environment An array of loudspeakers used to reproduce spatial sound fields, typically housed in an anechoic or soundproof room. In audiology, virtual sound environments have been used to simulate real-world soundscapes for the purpose of testing hearing aids.

List of Common Abbreviations

ABS Acrylonitrile Butadiene Styrene

AC(1) Autocorrelation at lag-1

ACI Acoustic Complexity Index

ADI Acoustic Diversity Index

AEve Acoustic Evenness

AWG American Wire Gauge

Bio Bioacoustic Index

CNN Convolutional Neural Network

CrI Credible Interval

ePTFE Expanded polytetrafluoroethylene

EWS Early Warning Signal

FLAC Free Lossless Audio Codec

H Acoustic Entropy

HARK Honda research institute Japan Audition for Robots with Kyoto University

IUCN International Union for Conservation of Nature

LiDAR Light Detection And Ranging

Loess Locally estimated scatterplot smoothing

Lowess Locally weighted scatterplot smoothing

M Median of the Amplitude Envelope

MAARU Multichannel Acoustic Autonomous Recording Unit

MEMS Micro Electro-Mechanical Systems

MUSIC MUltiple SIgnal Classification

NDSI Normalized Difference Soundscape Index

NDVI Normalized Difference Vegetation Index

PLA PolyLactic Acid

SPL Sound Pressure Level

$\boldsymbol{T_g}$ Glass transition temperature

VOD Vegetation Optical Depth

VODCA Vegetation Optical Depth Climate Archive

VSE Virtual Sound Environment

List of Figures

2.1 Diagrams illustrating the fold and transcritical bifurcation 43

3.1 Chapter 3 system overview and experimental structure 67

3.2 Field and virtual sound environment recording setups 69

3.3 Differences in spectral power between each VSE re-recording and the field recording for Sites 2 and 5 76

3.4 Plots of indices' average scaled like-for-like differences across all sites . . . 77

3.5 Field recording spectrogram, VSE recordings' spectral differences, and BirdNET predictions for 60 s in Site 2 and Site 5 recordings 79

3.6 Precision and recall of BirdNET predictions 81

3.7 HARKBird localisation error and BirdNET confidence against elevation values of bird calls added to VSE-based spatial natural soundscapes via ambisonic encoding or playback from individual loudspeakers 83

4.1 Depiction of MAARU's redesigned weatherproofing and UK and Malaysia deployments 95

4.2 MAARU array power patterns 102

4.3 MAARU 3D array power pattern for 4 kHz signals, with a 0° azimuth look direction 103

4.4 Histogram of BirdNET confidence scores for matching predictions in beamformed and omnidirectional recordings from the UK and Brazil 105

4.5 Number of BirdNET predictions and median species-specific average confidence for all recording conditions and sites 113

4.6 Number of BirdNET predictions (confidence > 0.25) and median prediction confidence of individual classes to either be attenuated or amplified via beamforming 114

4.7 Number of BirdNET predictions (confidence > 0.25) for most common species when beamforming to decompose habitat structure or distinguish neighbouring sites of distinct land use 115

4.8 Spectrograms showing Green Frog (*L. clamitans*) vocalisations in extract of omnidirectional and beamformed audio from US Site 2 116

4.9 Spectrograms showing road traffic and aircraft noise in extracts of omnidirectional and beamformed audio from UK Site 3 . . . 117

5.1 Overall satellite-derived vegetation resilience changes in Bornean rainforest 136

5.2 Spatial patterns of vegetation resilience changes in Bornean rainforest . . . 139

5.3 Relationship between vegetation resilience and potential influencing factors (sliding band size = 100 cells) . . . 140

5.4 Diurnal patterns and seasonal distributions of acoustic complexity and bioacoustic indices in recordings of the Amazon from Brazil Site 1 . . . 144

5.5 Early warning signals for acoustic complexity and bioacoustic indices from recordings of Brazilian Amazon . . . 145

B.1 Plots of scaled differences in indices' values between VSE and field recordings as a percentage of each index's field recording range, for all sites . . . 218

B.2 HARKBird localisation error against elevation values for avian calls embedded in VSE-replicated soundscapes by four different methods . . . 219

C.1 Version of Fig. 4.5 for UK Site 2 using three microphones for beamforming 220

C.2 Version of Fig. 4.7 showing remaining species for UK Site 2 . . . 221

C.3 Version of Fig. 4.7 showing remaining species for Malaysia Logged Forest 3 222

C.4 Version of Fig. 4.7 showing remaining species for US Site 2 . . . 223

C.5 Number of BirdNET predictions (confidence > 0.25) for all species in omnidirectional audio and audio beamformed with the subset of three microphones shown in Fig. 4.2 at UK Site 2 . . . 224

D.1 Relative time and impact of sensor shifts on time series of Vegetation Optical Depth and lag-1 autocorrelation . . . 229

D.2 Variance of Vegetation Optical Depth readings for Bornean rainforest since the early 1990s . . . 230

D.3 Satellite-sensed vegetation resilience change for Bornean rainforest with $\geq$ 60% evergreen broadleaf cover . . . 231

D.4 Satellite-sensed vegetation resilience change for Bornean rainforest with $\geq$ 90% evergreen broadleaf cover . . . 232

D.5 Robustness of Vegetation Optical Depth lag-1 autocorrelation to window length . . . 233

D.6 Vegetation resilience change of the Amazon rainforest . . . 234

D.7 Pairwise correlations between environmental factors . . . 235

D.8 Pairwise correlations between taxa . . . 236

D.9 Relationship between species richness of different animal taxa and resilience (sliding band size = 100 cells) . . . 237

D.10 Relationship between potential influencing factors and Bornean vegetation resilience when varying the band size to 50 cells 238

D.11 Relationship between species richness of different animal taxa and resilience when varying the band size to 50 cells 239

D.12 Recreation of Fig. 5.5, calculating early warning signals on one-week-long windows of acoustic indices advancing hourly 240

D.13 Recreation of Fig. 5.5, calculating early warning signals on four-week-long windows of acoustic indices advancing hourly 241

D.14 Recreation of Fig. 5.5, calculating early warning signals on acoustic indices for which data gaps were filled by linear interpolation 242

D.15 Recreation of Fig. 5.5 Site 1 on data resampled to 6 hr resolution 243

List of Tables

3.1 Average differences in spectral power per 1 kHz band between all VSE re-recordings and the reference field recordings . . . 75

3.2 Spearman's rank correlation coefficients and p-values assessing the relationship between MAARU's orientation and: (i) acoustic indices' and VGGish features' scaled like-for-like differences; (ii) HARKBird localisations; and, (iii) BirdNET confidence values . . . 78

3.3 Dunn post-hoc test results assessing for significant differences in BirdNET confidence and HARKBird localisations between all recording conditions . 80

3.4 BirdNET predictions and HARKBird localisations for bird calls added to spatial soundscape recordings from Silwood Park reproduced through our ambisonic VSE . . . 82

4.1 Results of Wilcoxon signed-rank tests applied to BirdNET confidence scores from beamformed and omnidirectional audio for individual species in the UK and Brazil . . . 106

4.2 Details of beamforming strategy at each recording site in UK, USA and Malaysia for Sub-Project 4 . . . 109

5.1 Early warning signals calculated on diurnal acoustic indices from recordings in Malaysian Borneo . . . 146

A.1 Expanded glossary of key concepts underpinning tipping point theory . . . 205

B.1 Breakdown of recordings used for Chapter 3 . . . 216

B.2 BirdNET results' statistics across all recording sites and conditions towards Chapter 3 Sub-Projects 1 and 2 . . . 217

D.1 Values of slope and breakpoint metrics of satellite-sensed vegetation resilience in Borneo and the Amazon . . . 225

D.2 Spearman's rank correlation coefficients (ρ) and associated p-values between resilience indicators, metrics and periods . . . 226

D.3 Spearman's rank correlation coefficients (ρ) and associated p-values between pre- and post-breakpoint values of Kendall's τ and slope for the spatially-averaged VOD AC(1) of sliding bands of grid cells advancing along gradients of various factors . . . 227

D.4 Results of robustness tests for early warning signals calculated on acoustic indices from recordings in Malaysian Borneo . . . 228

Chapter 1

Introduction: Into a world of sound

The Earth system is in peril [1–5]. Anthropogenic climate change, land-use change and resource exploitation are driving what many scientists regard as a Sixth Mass Extinction [6] and are pushing the planet to the brink. Numerous keystone ecosystems and fundamental parts of the climate system are approaching (or may have already passed) dangerous boundaries [7] for 'tipping points', where small change or disturbance could trigger collapse or abrupt transition to undesirable alternative states [2]. These nonlinear tipping phenomena are particularly worrying given their sudden nature and because they can be notoriously hard to reverse on human timescales. Since the initial submission of this thesis, the Global Tipping Points Report 2025 confirmed that we are now crossing the first of several major climate-related tipping points threatening the Earth system: the mass dieback of coral reefs from climate-change-induced ocean warming [1]. The outlook of our current trajectory is, therefore, existential for life on Earth.

However, averting the worst outcomes and restoring the balance in nature and climate is not beyond reach. Effective policy, management and collective action could yet safeguard a sustainable, liveable and prosperous future. But the action needed must add up to drastic, transformative change. Moreover, effective action depends upon a rigorous understanding of how natural systems operate, in what ways human activity impacts them and how they respond to protective and restorative interventions [8–11]. In the realm of ecosystem monitoring, autonomous sensing technologies are increasingly used to gain ecological insight over large scales, with less effort, bias and disturbance than traditional, manual surveys. Amongst these technologies, acoustic monitoring has emerged as a particularly capable tool for tracking biodiversity and ecosystem health at various scales,

owing to the wide array of sounds produced by animals and insects and their interactions with their environments [12–14].

The ecological importance of acoustics and the multifaceted value of sounds in nature have long been appreciated in different fields and cultures [15–17]. Sound is, of course, a key component of perception and communication in humans and other animals [17–20]. Furthermore, the sounds of environments – i.e., their *soundscapes* – impact human wellbeing [21–24] and are important to animal habitation and behaviour. For example, playing back the soundscapes of healthy environments in degraded ecosystems can help re-attract individuals and accelerate their recovery [25]. However, human activity has impacted soundscapes, either directly from the disruptive sounds of road traffic, other infrastructure and urban environments, or by habitat destruction and fragmentation, climate change and the large-scale eradication of biodiversity [26–28]. In Bolivia, for instance, the Chiguanco Thrush (*Turdus chiguanco*) has been found to vocalise less frequently, more loudly and at higher pitch in urban environments due to competing anthropogenic noise [29]. Anthropogenic biodiversity declines have also caused natural soundscapes to increasingly fall silent, which was noted as early as the 1960s in Rachel Carson's landmark text, *Silent Spring* [30]. In more recent years, there have even been calls for soundscapes themselves to be conserved as valuable natural resources of socio-ecological importance [15], and the ecological study of soundscapes has been formalised as the disciplines of 'soundscape ecology' [31] and 'ecoacoustics' [32].

Overall, soundscapes can be indicative of many features of ecosystems, spanning the sounds of individual animals to the layered symphony of soniferous biodiversity and acoustic signatures of natural habitat integrity. In this thesis, we explore new technologies and applications for the recording and analysis of natural soundscapes. We focus specifically on testing, developing and using multi-microphone recorders, which can capture information on where sounds are located in space, and investigate the potential for soundscapes to reveal indicators of ecosystem resilience loss that may signify impending tipping points. While we largely examine these two strands on the spatial and resilience-related features of natural soundscapes in parallel, we ultimately argue that sound-producing biodiversity is the bridge between them, as a key source of spatial heterogeneity in soundscapes and an ostensibly integral driver and product of ecosystem resilience.

In the following section, we provide some additional background on three key con-

cepts that are foundational to subsequent chapters. This is intended to complement the literature review in Chapter 2, which provides greater detail on tipping point theory and detection and the state-of-the-art in acoustic monitoring. The final sections of this chapter then highlight our Research Objectives and outline the structure of the thesis.

1.1 Key Concepts

1.1.1 Ecoacoustics: The Study of Soundscapes

While the study of animal communication and organism-specific sounds has existed for decades and is generally referred to as bioacoustics[1] [17, 33], ecoacoustics is a relatively new discipline focussed on analysing all the sounds in an environment that constitute its soundscape. A key part of ecoacoustics' roots lies in the work of R. Murray Schafer, who formalised the notion of 'the soundscape' in an academic context in the 1970s [34]. Schafer also proposed the discipline of 'acoustic ecology', which initially focussed on human perception and the impact of soundscapes on organisms [31, 34]. In the years after, key theories on the relationships between soundscapes and vocal animals were proposed, namely the Acoustic Adaptation Hypothesis [35] and the Acoustic Niche Hypothesis [16]. The former posits that the nature of animal vocalisations evolved to optimise their propagation in animals' chosen habitats, while the latter suggests that animals evolved to occupy unique frequency-time niches in soundscapes to avoid interfering with each other and other sounds. While empirical support for both hypotheses is mixed, they informed the development of various analytical techniques to monitor biodiversity from soundscapes, such as numerical summaries known as acoustic indices [36, 37]. In recent years, ecoacoustic analysis has also relied heavily on machine learning models that can predict the identity of species from recordings of their vocalisations and/or provide multidimensional numerical descriptions of ecologically-relevant soundscape features. More information on acoustic indices and machine learning models is presented in Section 2.2.3.

Alongside the development of ecoacoustic theory and analyses, technologies for recording natural soundscapes have in recent decades become smaller, cheaper and more capable. So-called 'autonomous recording units' have enabled acoustic monitoring to substantially increase in spatial and temporal scale with minimal human intervention. Some popu-

[1] Though we note the terminology can be slippery.

lar recorders, such as the 'AudioMoth' [38] or 'SwiftOne' [39] require occasional data collection and battery replacement (at up to months-long intervals, depending on the choice of recording schedule [38]). Others such as Sethi et al.'s custom-built recorder for the 'SAFE Acoustics' project [40, 41] and its subsequent iteration as the 'BUGG' device (bugg.xyz) leverage renewable powering (e.g., solar power) and network connectivity (e.g., 4G) to fully automate the collection and streaming of audio data in near real time. These advancements allow for 'passive acoustic monitoring' [12–14], capturing large amounts of natural soundscape recordings to closely track ecosystem health and biodiversity on hitherto impossible scales.

1.1.2 Spatial Audio and Multichannel Recording

Sound fields are three-dimensional and made up of acoustic pressure waves that travel outwards from sources at different locations in space. While recorders with a single microphone can capture the amalgamated soundscape, multi-microphone recording can provide spatial information on the direction or location of sound sources, typically based on differences in sounds' time of arrival at each microphone. Recording with microphone arrays can thus be highly valuable for tracking the distribution and position of soniferous biodiversity [42].

As with many innovations in passive acoustic monitoring, the use of multi-sensor acoustic recording to locate wildlife was pioneered in the marine realm, where visual imaging is challenging. For example, large distributed hydrophone arrays (i.e., with sensors tens of meters apart) have been used for the three-dimensional localisation of cetaceans since the 1970s [43]. In the terrestrial realm, distributed microphone arrays have been used to monitor various species including birds (e.g., [44]) and elephants (e.g., [45]), in order to gain insights into animal behaviour, movement, abundance and more (see [42, 46] for comprehensive reviews).

Distributed microphone arrays offer the benefit of determining the source location of target sounds, but face a major challenge from clock drift[2]. This introduces delays between recordings from individual microphones in the array and can thus cause localisation to fail [42]. Smaller, integrated arrays of multiple, closely-spaced microphones on a single

[2]Unless the microphones are connected to the same recorder via lengthy cables or synchronised via e.g., GPS.

multichannel device offer an alternative. Individual multichannel recorders cannot locate the exact position of sound sources, but they can use time delays to determine what direction sounds are emitted from. Crucially, they avoid clock drift, as the microphones share a common clock on the recorder. For these reasons – and as a result of innovations in other fields, such as multi-microphone hearing aids (e.g., [47]) – such integrated arrays have been increasingly adopted for direction of arrival estimation and directional filtering[3] in ecoacoustics [42].

Researchers have also developed custom multichannel recorders for ecoacoustic monitoring. For example, the 'VoxNet' system uses a four microphone array and features custom onboard processing software and wireless network connectivity for data streaming [48]. 'CARACAL' is a smaller and more recent low-power, low-cost recorder featuring a planar array of four miniature Micro Electro-Mechanical Systems (MEMS) microphones[4]. In this thesis, we focus on testing, refining and deploying the Multichannel Acoustic Autonomous Recording Unit (MAARU), a custom recorder that extends Sethi et al.'s fully-autonomous Raspberry-Pi-based system [40] with a six-microphone array [49]. Like CARACAL, MAARU's array features MEMS microphones and is planar. While this limits its capacity to estimate directions of arrival to two dimensions (the plane containing the microphones), many of the few low-cost microphone arrays in existence are planar (likely due to more complex three-dimensional arrays being more expensive to produce). Further, MAARU is to our knowledge the only fully-autonomous multichannel recorder for passive acoustic monitoring. Therefore, we view MAARU as a prime multichannel recorder to study and deploy towards advancing the field of spatial ecoacoustics.

In parallel to hardware progression, various analytical techniques have been developed for estimating sound-source location and/or direction of arrival – including numerous machine learning approaches in more recent years [50] – and applied for tracking terrestrial [42] and aquatic wildlife (e.g., [51, 52]). Additionally, several tools have been created that implement localisation and direction of arrival estimation algorithms for bioacoustics research, such as the 'Where's Whaledo' *MATLAB* package [52] and 'HARKBird' [53, 54]. The latter is a freely-available program for estimating the direction of arrival of bird calls in multichannel recordings made with integrated microphone arrays. It is based on the robot audition software, HARK (Honda research institute Japan Audition for Robots

[3] Further explained in Section 4.1.

[4] Since extended to eight microphones: https://github.com/OpenWild/CaracalEvo_HW

with Kyoto University) and uses the MUSIC (MUltiple SIgnal Classification) algorithm for direction of arrival estimation. MUSIC is a widespread eigenspace-based method for estimating the directions of arrival of multiple target signals in multichannel recordings, developed in 1986 [55]. Importantly, the MUSIC algorithm assumes that there are fewer target signals than microphones (or more generally, sensors) in the recording array, and that the target signals are uncorrelated.

While innovations in spatial ecoacoustic monitoring have been advancing with novel hardware like MAARU and software like HARKBird, methods to test these technologies lag behind other fields. In audiology, for instance, various sophisticated approaches have been developed to allow for the precise testing of hearing aids in controlled environments (e.g., [56, 57]). In general, the testing of ecoacoustic hardware has tended to focus on field testing and experiments that involve playing back test sounds from loudspeakers [58]. However, such methods often suffer from poorly documented protocols [58] and will be influenced by natural variability in field conditions (weather, variation in vocal activity of species present, etc.). We thus note that improving methods for testing spatial ecoacoustic tools with more precise and controlled techniques could be very beneficial towards benchmarking their performance.

1.1.3 Feedbacks Govern the World

While tipping points are explained in much more detail in the following chapter, *feedbacks* underpin them and are the third and final concept that merits further background here. In the general sense, feedbacks refer to the outputs of a system 'feeding back' into its inputs, either to correct course via negative feedback, or to self-amplify via positive feedback. Sometimes, the term also denotes the outputs of processes or systems feeding into others. Feedbacks are integral to dynamical systems across domains. Perhaps one of the best-known feedbacks occurs in acoustic amplification systems: the characteristic unpleasant squeal of a microphone picking up its amplified signal from a nearby loudspeaker. In engineering, feedbacks are critical to control systems with applications from thermostats to robots, while in ecology feedback loops are key features of homeostasis and nutrient recycling.

More generally, in many systems, including the climate and ecosystems, feedbacks also play a key role in abating or driving transitions between different states, such as

different stable animal population densities or alterative stable levels of biomass in forests [59, 60]. Negative feedbacks maintain stable states by counteracting deviations in the system caused by internal fluctuation or external perturbation. Positive feedbacks amplify such deviations, and can drive runaway change towards a new stable state. Tipping points can lead to such transitions, and in climate and ecological systems, they could have dangerous consequences such as ecological collapse and drastic shifts in weather. These transitions may also feed into other parts of the Earth system, raising the risk of possible 'tipping cascades' [61].

As a concrete example, there are widespread concerns that tropical forests could tip into savannah-like or treeless states through events of mass tree mortality or 'dieback' [62–64]. One of the main mechanisms that could drive such a transition is the breakdown of moisture recycling feedbacks in response to drier and warmer conditions, as well as forest loss from fire and deforestation [63, 65]. The characteristic high humidity and vegetation density of tropical rainforests are maintained by high precipitation. In the Amazon rainforest in particular, this precipitation is supplied by incoming moisture in upwind areas and cascades of moisture recycling via evapotranspiration towards and across downwind areas [65]. Reduced precipitation coupled with forest loss produces a drier environment that is more conducive to the spread of fire and reduces the moisture available for recycling to downwind areas [63]. The stabilising effect of moisture can thus succumb to self-amplifying positive feedbacks of drier conditions, fire and dieback [63, 66]. Should this happen for the Amazon, it will release a vast amount of additional carbon dioxide into the atmosphere and eradicate the cooling effect of the forest's evapotranspiration [63,67], thereby driving further global warming [2]. It is therefore clear that understanding feedbacks in natural systems is critical for evaluating whether and how they might tip and what broader consequences their tipping points might have.

1.2 Research Objectives

With the aim of developing methods to assess the spatial and resilience-related features of natural soundscapes, we set three core Objectives for the research in this thesis:

1. To develop a representative, controlled and repeatable means of assessing spatial ecoacoustic technologies, including multichannel recorders, in laboratory settings.

2. To determine whether multichannel audio processing for directional filtering and amplification ('beamforming') could enhance biodiversity monitoring.

3. To explore whether acoustic monitoring could reveal Early Warning Signals (EWSs) of ecosystem tipping points and support satellite remote sensing in assessing tropical rainforest resilience.

1.3 Thesis Outline

In line with our Research Objectives, this thesis tackles challenges and questions across the full ecoacoustic monitoring pipeline.

We begin by assessing theory and prior research with a literature review in **Chapter 2**, which we frame as a perspectives piece advocating for the potential of acoustic monitoring to anticipate ecosystem tipping points. In many cases, tipping points are preceded by declines in resilience, such as slowing recovery from perturbations. Numerous tipping point EWSs based on these phenomena have been developed and identified in natural systems, and we propose a methodological framework for applying such EWSs to acoustically-derived ecological time series. We also emphasise practical and theoretical barriers to realising this framework, including challenges with gathering or assimilating acoustic data at sufficient scale to detect meaningful EWSs of resilience loss driven by global climate and environmental change.

Fieldwork in the tropical rainforests of Malaysian Borneo was planned from the outset as a key component of research towards Objectives 2 and 3 of this thesis. However, it was initially not viable due to ongoing COVID-19-related restrictions in Malaysia. We therefore turned our attention to the start of the ecoacoustic monitoring pipeline and, in line with Objective 1, developed and validated a novel approach to testing spatial ecoacoustic technologies, which we present in **Chapter 3**. Inspired by how multichannel hearing aids are tested in audiology, we built a 'Virtual Sound Environment' (VSE) with a spherical array of 25 loudspeakers to simulate spatial natural soundscapes as a controlled and life-like setting for testing ecoacoustic recording technologies. We assess how well the VSE simulates natural soundscapes, and then pilot using it to evaluate ecoacoustic hardware and software tools, including MAARU. We find limitations to the ambisonic soundscape reconstruction technique we used, but are able to gain insights into the performance of

the various technologies we test, including the impact of microphone array orientation on bird call direction of arrival estimates with MAARU.

While multichannel recording shows promise to enhance ecoacoustic monitoring, few multichannel recorders have been designed for long-term, autonomous deployments. In **Chapter 4**, we redesign MAARU's weatherproofing to be more robust to prolonged environmental exposure through a process of iterative refinement and field testing. Towards Objective 2, we then explore how beamforming with MAARU to amplify sounds from target directions can enhance bird call classification in temperate and tropical environments. As **Chapter 3** revealed limits to the accuracy of MAARU's direction of arrival estimates for sound sources outside its microphone plane, we deploy MAARUs in the UK, USA, and Malaysia in orientations that allow for fixed-direction beamforming to target ecologically relevant sound sources or directions. We find that this simple but scalable approach has potential to reveal habitat and landscape structure and detect the spatial heterogeneity of biodiversity.

In **Chapter 5**, we return to Objective 3 and the question of resilience. Ecoacoustics has yet to be leveraged to detect EWSs of tipping points, but global satellite imagery has proven particularly effective for doing so, including to track vegetation resilience in the tropics. Given further fieldwork delays and recorder hardware failures in Borneo, **Chapter 5** therefore focusses first and foremost on extending the detailed investigation of rainforest vegetation resilience using satellite data to Borneo. We find increases in EWSs of a vegetation tipping point across a period of over 20 years, with an acceleration of resilience loss from the early 2000s, as has been observed in the Amazon. We relate spatiotemporal patterns in this declining resilience to potential driving factors including climate and human activity as well as animal biodiversity, which has been underexplored in relation to systemic vegetation resilience. We then seek to complement these findings by searching for EWSs in acoustic descriptors of soniferous biodiversity from recordings along a land-use gradient in Malaysian Borneo and audio data gathered by colleagues over a two-year period in the Amazon. Despite our attempt to realise the framework we outlined in **Chapter 2**, we find the scale of our data too limited to draw conclusive insights into whether soniferous species in rainforests are also losing resilience. However, we also identify preliminary evidence that EWSs in the acoustic domain might reveal changes in how fast soniferous biodiversity recovers from perturbation, with potentially

slower responses in logged forest. We thus posit that multimodal monitoring of rainforest resilience with satellite and acoustic data merits further investigation with longer-term recordings. Such investigation could further probe the resilience of sound-producing biodiversity and its relationship to tipping points of rainforest dieback that threaten local species, communities and the wider Earth system.

Chapter 2

Literature Review: Prospects for acoustically monitoring ecosystem tipping points

Summary

Many ecosystems can undergo important qualitative changes, including sudden transitions to alternative stable states, in response to perturbations or increments in conditions. Such 'tipping points' are often preceded by declines in aspects of ecosystem resilience, namely the capacity to recover from perturbations, that leave various spatial and temporal signatures. These so-called Early Warning Signals (EWSs) have been used to anticipate transitions in diverse real systems, but many of the high-throughput, autonomous monitoring technologies that are transforming ecology have yet to be fully leveraged to this end. Acoustic monitoring in particular is a powerful tool for quantifying biodiversity, tracking ecosystem health, and facilitating conservation. By deploying acoustic recorders in diverse environments, researchers have gained insights from the calls and behaviour of individual species to higher-level soundscape features that describe habitat quality and even predict species occurrence. Here, we draw on theory and practice to advocate for using acoustics to probe ecosystem resilience and identify emerging and established EWSs of tipping points. With a focus on pragmatic considerations, we emphasise that de-

spite limits to tipping point theory and the current scale and transferability of data, acoustics could be instrumental in understanding resilience and tipping potential across distinct ecosystems and scales.

2.1 Introduction

Across diverse domains and scales, dynamical systems are known to possess multiple stable states between which they may transition [68, 69]. Beams buckle, populations explode or collapse, financial markets crash, human brains experience seizures or different psychological states, ice ages begin and end, and entire ecosystems small and vast can transform drastically faster than expected. Such transitions, sometimes termed 'regime shifts' [70–73] or 'critical transitions' [68, 74], can arise due to various underlying mechanisms that may produce gradual, fast, or sudden, nonlinear shifts. 'Tipping points' denote those caused by small perturbations amplified by positive feedback [59, 60], extending the term popularised by Gladwell [75] (though its use has varied somewhat in the past [74, 76]). Tipping points produce abrupt, qualitative change to the state of a system that may be difficult or impossible to reverse [59, 60]. As such, they have been widely researched in recent decades, especially in climate science [2, 59, 60, 77], ecology [62–64, 70, 71, 78–81] and their intersection in Earth system science [82–86], where the risk to natural systems and humanity is particularly great [61]. Despite the urgent need to monitor and anticipate tipping points, many advancing autonomous monitoring technologies that empower data-driven ecology have not yet been applied to real-time assessment of impending ecosystem tipping points. Here, we spotlight acoustic monitoring as a particularly insightful technology to which existing tipping point EWSs could be readily extended for enhanced understanding of whether and how ecosystems under intensifying environmental change worldwide may tip.

Tipping points are predominantly brought about by bifurcations (Fig. 2.1; Section 2.1.1), noise that is internal or due to stochastic forcing pushing the system out of its current stable state's 'basin of attraction' (Fig. 2.1; Section 2.1.1), or, sometimes, rapid forcing in excess of a threshold rate [60, 79, 87]. Bifurcations are a sudden qualitative change to the dynamics of a system – typically, the loss of the current state's stability, resulting in a shift to an alternative stable state – in response to gradual change in the

control parameter(s) of the system passing some critical threshold (a.k.a. the bifurcation point) [60, 69, 79]. Much of the work on EWSs for tipping points has focussed on bifurcations as they are widespread and satisfy universal mathematical laws that can be detected in temporal and spatial data [74].

Moreover, according to the centre manifold theorem, the dynamics of complex systems near a local bifurcation simplify to a subset of lower-dimensional, prototypical 'normal forms' [69, 88], promoting a strong basis for detecting generic EWSs in real systems. These EWSs generally exploit the declining stability before a bifurcation, which causes critical slowing down of a system's response to perturbation [69, 74, 89], yielding increases in variance and autocorrelation – two of the most widespread EWSs. The capacity to recover from perturbations is often also considered integral to a system's resilience – though this has long been discussed in ecology [90–95]. In addition, noise-induced tipping and bifurcations often work together in reality, since the weakening of a current state's stability (negative feedback; Section 2.1.1) towards a bifurcation increases the chances that noise or a typical perturbation can push the system to an alternate attractor – sometimes with devastating consequences. EWSs can thus indicate increased risk of tipping as a system loses (aspects of) resilience, spurring prolific research into their development and application [60].

In ecology, the notion that dynamical natural systems from populations to whole ecosystems may possess multiple stable states has long existed [96]. The concept of non-linear state transitions and their thresholds gained traction from the 1960s [90, 97, 98] after they were described in simple ecological models, namely of grazing [99] and harvesting systems [96, 98], predator-prey interactions [100], and insect outbreaks [98]. The observation of large-scale, abrupt regime shifts in various real ecosystems – including clear lakes becoming turbid, macroalgae dominating coral reefs, forests switching to grasslands, and arid environment desertification – also paved the way for further research on characterising and anticipating such transitions [78]. Since early work suggested critical slowing down could be measured in fluctuations of the system state caused by stochastic forcing [101, 102] and models revealed that critical slowing down increases autocorrelation [102, 103] and variance [104], these metrics and other EWSs have been identified in a fast-growing number of systems (see reviews in [60, 81, 105]).

A key challenge in detecting EWSs in real ecosystems is the limited data of sufficient

spatial and temporal scale, quality, and resolution [105, 106]. While EWSs of past climatic shifts have been demonstrated in paleoclimate data [107, 108], historical data for ecosystems is less forthcoming. One solution is to study various spatial patterns, such as spatial variance and autocorrelation [109], or vegetation patch size and distribution (relevant to e.g., self-organisation of semi-arid vegetation under a drying climate) [80]. Another is space-for-time substitution: monitoring ecosystems along spatial gradients of human, environmental, and/or climatic pressures [64, 110–112]. In the past 15 years, satellite remote sensing has also emerged as a vital tool for ecosystem monitoring as it provides exceptional resolution and coverage of the Earth, and certain sensors now have 50-yr records [85]. Remote sensing has thus become the predominant data source for ecosystem tipping point research [60] and has been leveraged to identify temporal and/or spatial EWSs in various ecosystems globally, including the Amazon [62], other tropical rainforests [64, 113] and even the biosphere at large [84].

The availability of ecological field survey data has, however, grown significantly in recent years, thanks to the emergence of various high-throughput, autonomous monitoring technologies [8], shaping the view that ecology is increasingly a "big data science" [114]. These technologies include acoustic sensors, environmental DNA sensing, camera traps and other image/electromagnetic-based systems such as thermal, infrared and Light Detection And Ranging (LiDAR) sensors [8, 115]. Numerous studies have leveraged data from these sensing modalities, but we argue there remain considerable untapped opportunities to search within large and expanding monitoring datasets for EWSs and integrate them, alongside satellite sensors, into multimodal, holistic models of ecosystems for more comprehensive assessment of ecosystem health, dynamics, resilience, and proximity to tipping.

We contend that ecoacoustics – the analysis of natural soundscapes – presents a particularly promising avenue for monitoring ecosystem tipping points, for several overarching reasons. First, it is information rich. As sound propagates from an emitter in all directions, past many barriers and over fairly large areas, acoustic monitoring can capture information outside the field of view or spatial reach of other sensors. Second, it is relatively inexpensive, particularly in terms of the declining cost of recording hardware. Third, it provides a high temporal resolution record for examining ecosystems at many levels, from individuals to populations and higher-level metrics of overall ecosys-

tem health [14, 32]. By deploying acoustic autonomous recording units, researchers have undertaken the passive acoustic monitoring of diverse ecosystems, uncovering a wealth of ecological, behavioural, and conservation-relevant insights across the globe [116] on larger scales and with less bias and disturbance than manual field surveys [12, 13, 45]. Passive acoustic monitoring has been used to study fish [117], cetaceans [118], frogs [119], insects [120], birds [121], terrestrial mammals [45, 122] and facilitate estimates of overall biodiversity and habitat quality [123]. Larger-scale passive acoustic monitoring deployments are also increasing, capturing recordings over landscapes [124, 125], countries [126] and continents [127]. Further, passive acoustic monitoring has captured aspects of stability and resilience by revealing various ecosystem responses to disturbances including human infrastructure development [128], forest fragmentation [129], species invasion [130], fires [131], and typhoons/hurricanes [132, 133]. Passive acoustic monitoring is now well-poised to build on its successes in tracking ecosystem health and biodiversity to become a pillar of resilience and tipping point monitoring.

In this chapter, we review a broad, interdisciplinary literature to outline the prospects for passive acoustic monitoring of ecosystem tipping points from a theoretical and practical point of view, including considerations for identifying EWSs in acoustic data. Throughout, we focus primarily on critical slowing down-based EWSs, given their ongoing dominance in tipping point research, but consider others where appropriate. We close by discussing the key challenges and opportunities for using passive acoustic monitoring to identify EWSs.

2.1.1 Box 1: Basic Theory of Bifurcation Tipping Points

Dynamical systems are characterised by a state, x, that changes over time according to a growth function $f(x, B)$ dependent on x's current value and some parameter(s) B:

$$\frac{dx}{dt} = f(x, B) \tag{2.1}$$

Real natural systems are often modelled as stochastic (with a noise term) to capture their continual fluctuation (see Table A.1). Classic ecological examples of x are species abundance (population dynamics) or vegetation biomass.

The system will often possess equilibria (steady states): values of $x = x^*$ such that

$f(x^*, B) = 0$. In single-variable systems, these equilibria are stable states (attractors) when λ, the local gradient $\frac{df(x,B)}{dx}\Big|_{x=x^*}$ is negative, corresponding to restorative negative feedback (Fig. 2.1c,d). The magnitude of λ is inversely proportional to the system's recovery time from small perturbations [89]. In higher dimensions, stability is determined using the real part of the dominant eigenvalue (see Table A.1).

Various bifurcation normal forms common to natural systems are brought about by gradual forcing of parameter(s) B that causes λ to tend to zero (Fig. 2.1c,d). Consequently, stability is progressively lost and the system recovers more slowly from perturbations – so-called 'critical slowing down'. This results in the system displaying longer and larger fluctuations about its current attractor state, such that both autocorrelation (the system's 'memory') and variance increase prior to the bifurcation point, at which $\lambda = 0$ (derivations in Section A.1).

While resilience has been much debated in ecology, Holling's definitions of 'engineering resilience' – the recovery rate from perturbation – and 'ecological resilience' – the amount of disturbance a system can withstand without changing its function – remain prevalent [90, 91]. Both are captured by the 'potential landscape' of the system:

$$V(x) = -\int f(x, B)dx \tag{2.2}$$

Here, stable states are valleys, often termed 'basins of attraction' [97], separated by peaks from unstable equilibria (Fig. 2.1g). In the popular 'ball and cup' analogy of resilience, the system is a ball rolling inside a cup corresponding to the system's current basin of attraction. The ball is always moving due to noise and perturbations, but typically rolls back to the bottom at a rate depending on the steepness of the cup's walls. The steeper, deeper and wider the cup, the more 'resilient' the system.

As $V''(x) = -\lambda$, the cup becomes gentler and shallower as λ approaches zero. Noise or typical perturbations can thus more easily push the system to an alternative attractor (Fig. 2.1g), should one exist. Hence, while ecological resilience may depend on more than just the potential landscape (e.g., ecosystem reorganisation), both engineering and ecological resilience are linked given their dependence on λ.

Which bifurcation normal form a system will undergo depends on the nature of the system dynamics. Fig. 2.1 illustrates two widespread cases in ecology using the archetypal example of a single-species population being harvested at an increasing rate (further

explained in [98, 105]): the gradual shift to an alternative stable state (here, extinction) via a transcritical bifurcation (left column), and the sudden, discontinuous jump to an alternative stable attractor via a fold (a.k.a. 'saddle-node' or 'catastrophic') bifurcation, which typifies tipping points with a large, abrupt change that is hard to reverse due to hysteresis. These dynamics could be reflected in measurements, proxies or correlates of the system state, which could in turn be used to detect EWSs of bifurcation tipping points. Section 2.2 outlines the step-by-step process for doing so and a brief applied example is provided in Section 2.2.6 (Box 2).

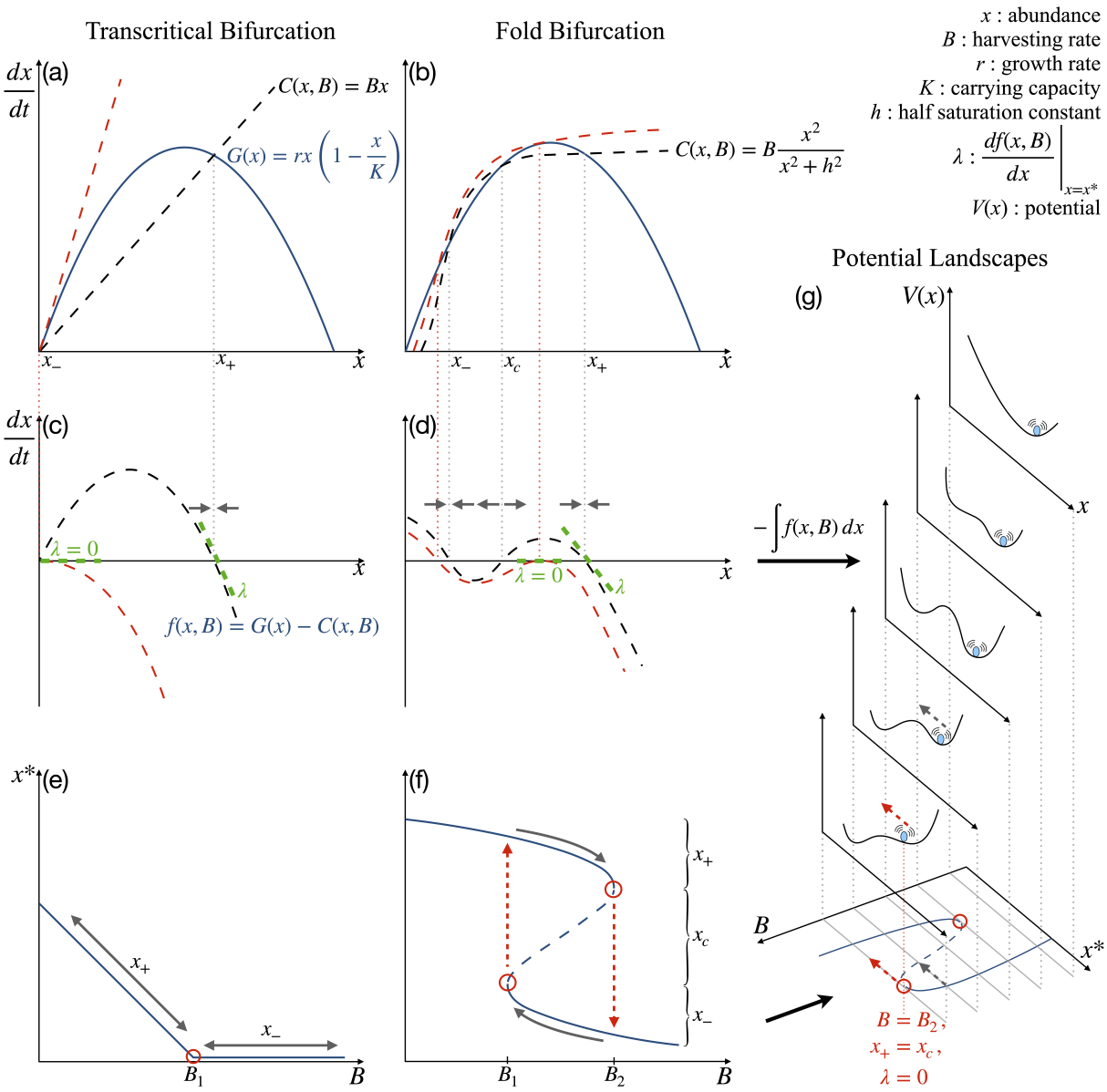


Figure 2.1: Graphs depicting the transcritical and fold bifurcation, illustrated with a single-species population model with harvesting (details in [98, 105]). Panels **a** and **b** show the population growth rate function $G(x)$ (blue line) and harvesting rate functions $C(x, B)$ (dashed lines). In panel **a** the harvesting increases linearly with abundance x, whereas in panel **b** it saturates. Black dashed lines depict a lower harvesting rate, B, resulting in multiple equilibria (net zero growth rate) where the harvesting and growth rate functions intersect. Red dashed lines depict harvesting at a higher critical rate at which certain equilibria disappear – the bifurcation point. Panels **c** and **d** show the net growth functions, $f(x, B)$, colour-matched to the two harvesting scenarios in panels **a** and **b**. The net growth functions possess equilibria where these curves cross the x-axis at x_-, x_c and x_+. These equilibria are stable when their gradient, λ, is negative (corresponding to negative feedback). At the bifurcation point (red dashed lines): in panel **c**, the population has been gradually driven to extinction and the point at zero abundance is now a stable equilibrium; in panel **d**, the stable equilibrium x_+ and unstable equilibrium x_c collide and the system jumps to its lower stable equilibrium x_-. In both cases, λ tends to zero (critical slowing down, see Section 2.1.1). Panels **e** and **f** show bifurcation diagrams, in

which arrows indicate how the equilibrium states of the systems vary with the parameter B (here, the harvesting rate). Red circles mark bifurcation points and red arrows indicate discontinuous transitions. Note a large decrease in B is required to bring x^*_- back to x^*_+ due to hysteresis. Panel **g** illustrates potential landscapes (see Section 2.1.1) at different values of B along the fold bifurcation diagram (adapted from (Scheffer et al., 2001)), in which peaks represent unstable equilibria and valleys stable equilibria. Following the 'ball and cup' analogy (Section 2.1.1), the system is depicted as a ball constantly moving from noise and perturbations in its current valley (a.k.a. basin of attraction), which gets shallower as λ approaches zero, corresponding to a loss of resilience (Section 2.1.1). The arrows here either indicate the system moving to its alternate attractor due to the bifurcation itself (red arrows) or sufficiently large noise or a perturbation (grey arrows).

2.2 Step-by-Step Considerations for Acoustically Identifying Tipping Points

2.2.1 Step 1: Will it Tip? Identifying Tipping Potential

While critical slowing down has proven an effective basis for warning of tipping points in a range of systems [60, 74], it can occur without a tipping point and certain tipping points can occur without critical slowing down [105, 134, 135]. Thus, determining whether the study system has alternative states that it may tip between is a key first step [85, 134]. Understanding the type of transition that can occur, its underlying mechanics, and which variables in a multivariate system are likely to tip and/or show critical slowing down can inform an appropriate strategy for identifying EWSs. Although doing so can be challenging and may bias the perceived effectiveness of EWSs [60], it is necessary to avoid false alarms, especially given the novelty of using acoustics to anticipate tipping points.

Whether and how an ecosystem may tip can be uncovered in various ways. Theory, models and an understanding of the various feedbacks that maintain stability or drive state transitions provide one basis. For instance, removing a top predator may cause populations at lower trophic levels to explode [78] while adding predators may drastically alter food web structure [71]. Alternatively, rainforest dieback or deforestation may cause drier local conditions that spur fires and reduce moisture recycling driving further dieback, ultimately tipping to a grassy or treeless state [63, 65]. Historical records (e.g., paleoclimate data [107]) and/or other sensing modalities (e.g., satellite data [85]) may also indicate EWSs or evidence of previous regime shifts in the study system (or similar systems). For example, decades of manual point counts were recently used to reveal a

tipping point in an at-risk bird community [136].

Where existing data reveal alternative stable states across space or time, 'potential analysis' can empirically estimate the potential landscape based on the probability distribution of the system state in relation to a key driving parameter (e.g., temperature or precipitation) [64, 108]. This method has been applied to tropical forests, where remotely sensed tree-cover estimates revealed precipitation thresholds for bifurcations to savannah or treeless states [64]. If the study system is impacted by other established tipping systems, tipping in those systems could potentially propagate [85, 134] or cascade [82] into the study system. Whether the study system would show EWSs, however, depends how the systems interact [60].

Understanding multivariate systems is more challenging, as it is often unclear how variables will respond to forcing. EWSs may be inconsistent across variables and critical slowing down could be less pronounced due to the dominant eigenvector's diminishing importance in a high-dimensional system with many eigenvectors [60, 105, 137]. However, various approaches to anticipating transitions in multivariate systems have emerged, such as network analysis and dimensionality reduction techniques [60]. Thus, prior to acoustically detecting EWSs in a multivariate system, it might be especially important to leverage existing insights into whether it can tip and which key variables or indicator species [138] might best reflect that possibility.

2.2.2 Step 2: Gathering Data at Sufficient Scale and Resolution

Bifurcations are predicated on the gradual, predominantly monotonic external forcing of the system on a slower timescale than its inherent response to perturbations [60]. Detecting critical slowing down requires monitoring long enough to capture forcing while sampling the system state more frequently than its response rate [85]. For example, Amazon rainforest vegetation was estimated to have a characteristic recovery time of approximately 1.2 months [62] (this arises from the time for leaf regrowth) and faces forcing from human activity and climate change (declining precipitation) on the scale of years to decades. Hence, Boulton et al. [62] assessed EWSs of Amazonian rainforest with over 20 years' remote-sensed vegetation data of monthly temporal resolution.

Theory and such existing data can indicate the required monitoring duration and resolution (e.g., [139]). Passive acoustic monitoring offers the advantage of very high

(sub-millisecond) temporal resolution and autonomous recording units can record over years (e.g., [45, 118]), though decadal-scale passive acoustic monitoring is currently rare. Space-for-time substitution may allow for shorter timescales by monitoring along spatial gradients to forecast the fate of a progressively stressed ecosystem [110–112]; monitoring along land-use or disturbance gradients is already relatively common in passive acoustic monitoring (e.g., [41, 133, 140, 141]).

The required spatial extent of monitoring depends on the scale of the study ecosystem. Autonomous recording units typically detect sounds up to 10s to 100s of meters away, though this strongly depends on sources' sound pressure level and environmental factors (e.g., weather, vegetation density) [142]. Thus, while the full extent of vast ecosystems of up to sub-continental scale, like the Amazon, can be readily monitored by satellites, doing so with acoustics presents a formidable challenge due to the cost and effort of such large deployments. However, large-scale passive acoustic monitoring deployments are feasible (e.g., [125–127]) and can even be achieved through citizen science (e.g., [143]). Where only smaller deployments are possible, identifying representative subregions offers one solution, including relevant spatial pressure/disturbance gradients. Monitoring replicates of such gradients or multiple sites at each pressure level can validate the robustness of such methods.

2.2.3 Step 3: Extracting Relevant Ecological Metrics from Acoustic Data

Methods for extracting ecological information from acoustic data are numerous and may capture ecosystem state variable(s) to monitor for EWSs. Most utilise the spectral properties of sound – the amount of energy at different sonic frequencies – usually calculated by applying a Fourier transform to the time-domain acoustic signal [12, 13]. The frequency distribution is typically represented with spectrograms, 2D image-like plots of the amplitude or power (colour value/intensity) of different frequency bands (y-axis) over time (x-axis). Broadly, different taxa typically vocalise or otherwise produce sound in specific and sometimes distinct frequency ranges [31, 144], and different species' calls possess unique spectral and temporal signatures, allowing them to be identified by various means, from manual inspection and simple thresholding to cross-correlation and more advanced machine learning classification [12, 13, 145, 146]. Spectral properties are also often

inspected to verify the performance of downstream analyses and tools (e.g., [145, 147]) as part of so-called 'sound truthing' [37, 148].

While manual methods were once dominant, the past decade has seen a drive towards faster automated techniques [12, 13, 145], namely acoustic indices, species call classifiers and machine learning model embeddings. These methods will be fundamental to efficiently derive ecological information at the scale required for identifying EWSs.

Acoustic indices are 'human-derived' numerical summaries of acoustic data designed to characterise the biophony (sound of organisms), anthropophony (human sound) and geophony (geophysical / environmental sound, e.g., wind and rain) in soundscapes based on various spectral and temporal features [37, 149, 150]. To-date, over 60 have been developed [151]. Prevalent examples include the acoustic complexity index, bioacoustic index and the normalized difference soundscape index. The acoustic complexity index assesses the total amount of temporal variability in different frequency bands as a proxy for avian vocal activity or other irregular biophony and geophony [152]. Specifically, it computes the sum of differences between successive linear amplitude values within short temporal clusters in each frequency band of a spectrogram. It then divides these sums of differences by the sum of the values for their respective clusters, and takes the total of such calculations for all time-frequency clusters across the spectrogram [149, 152]. As the index is cumulative, averaging it over longer recordings (e.g., per-minute mean) is advised [149, 153]. The bioacoustic index [154] is related to the total amount of intensity in a range of frequencies (typically 2-8 kHz) above the quietest frequency band [153] and was originally intended to capture avian abundance in Hawaii. It is calculated by taking the mean amplitude (in decibels) of each spectrogram frequency band within the preset frequency range, subtracting the minimum of those values from them, and then taking their sum divided by the Nyquist frequency (half the sampling frequency) [149]. The normalised difference soundscape index [155] seeks to characterise human disturbance on soundscapes by comparing the amount of anthropophony and biophony in recordings. It splits recordings into bands between 1-2 kHz assumed to be associated with anthropophony and bands between 2-8 kHz assumed to be associated with biophony, and then computes the ratio between the difference and sum of the power spectral density of the largest biophony and anthropophony bands [153, 155]. Higher values are thus intended to indicate more dominant biophony.

Such indices can describe soundscapes and serve as proxies for biodiversity [13,37]. For example, avian diversity has been found to positively associate with the bioacoustic [144, 154, 156, 157] and acoustic complexity [156–158] indices. However, negative correlations between the acoustic complexity index and avian diversity have also been reported [144]. Indeed, recent work has cautioned the use of acoustic indices for estimating biodiversity due to inconsistencies in their relationship to biodiversity between regions and taxa [36, 159]. Hence, careful study design and local validation to guide the interpretation of acoustic indices is advised [37, 144].

Species call classifiers use the time-frequency characteristics of different species' calls to identify vocalising species [12, 146], offering a way to directly measure biodiversity and track the presence, distribution, and behaviour of particular species (e.g., through occupancy models [125, 160]). Current state-of-the-art classifiers use a range of different deep learning models and architectures, in particular various forms of Convolutional Neural Network (CNN), be they custom-designed or pre-existing architectures such as ResNet [161], VGG [162], VGGish [163], and EfficientNet [164]. Deep learning classifiers have been developed for many taxa, including birds, cetaceans, bats, elephants, primates, frogs, fish, and insects (examples in [146]). Classifiers may be species-specific or able to identify a range of species within one or more taxa, and are typically trained on large datasets (often at least thousands of example calls) under a supervised approach, requiring the often-laborious labelling of training data (though see [165] on unsupervised approaches).

In recent years there have thus been efforts to consolidate datasets and models towards capable, 'off-the-shelf' classifiers that are easy to access and use. Examples include the BTO Acoustic Pipeline, which brings together classifiers for various birds, small mammals, insects and bats (e.g., [166]), and BirdNET [167], a CNN capable of classifying over 6,000 avian species globally. BirdNET was trained on a substantial quantity of data, including 100,000s recordings from public wildlife sound repositories, Xeno-Canto and the Macaulay Library. It was initially trained to classify North American and European bird species, but has since been expanded to global coverage and can now classify certain non-avian species, such as anurans [168], as well as 11 'non-event' classes (e.g., human voices, dogs, engines, and guns). A limitation of BirdNET's training data being heavily skewed to large public sound repositories is that these datasets primarily contain focal recordings of

wildlife (i.e., made with directional microphones and/or close to vocal individuals), which may not be reflective of actual passive acoustic monitoring conditions. In reality, passive acoustic recordings of an amalgamated soundscape made with autonomous recording units featuring a single omnidirectional microphone will contain many overlapping callers across a wide range of distances to the recorder and more background noise than much of BirdNET's training data. This presents a 'domain shift' which may hinder BirdNET's performance on real-world passive acoustic monitoring data [169–171]. Nonetheless, BirdNET has become one of the most widely-used species call classifiers, unlocking insights into avian species and communities around the world [170, 172].

As well as facilitating many ecological studies, pre-existing models like BirdNET are also increasingly being retrained to detect further species. This approach, termed 'transfer learning' [173], leverages the existing learning of models to learn new classes efficiently, sometimes with very little training data (e.g., < 10 examples) – so-called 'few-shot learning' – hence it is particularly pertinent for rare/endangered species with scarce recordings [174, 175]. Transfer learning often leverages existing models' learning by using deep learning model embeddings [173, 174]. These capture salient features of the input data and are obtained by passing audio into a pre-trained deep learning classifier and taking the output of an intermediate (usually, penultimate) model layer. This output is a vector of sometimes hundreds or thousands of numbers that characterises the input recording (sometimes termed a 'feature set' [123] or 'fingerprint' [176]). The embedding vector can be thought of as an abstract, machine-learned, multivariate acoustic index that can describe ecologically-relevant patterns in a soundscape and potentially provide proxies for biodiversity [123, 159, 174, 177]. As with traditional, human-derived acoustic indices though, model embeddings' relationships to biodiversity can be inconsistent across recording conditions (different regions, recording hardware, etc.), hence they too should be used with caution and careful calibration [159].

Nonetheless, embeddings from the general-purpose audio classifier VGGish [163] have outperformed acoustic indices in various downstream habitat quality and biodiversity classification tasks [123,176], are more robust than acoustic indices to audio compression [176], and have supported various other biodiversity assessments (e.g., [178, 179]). Performance in various ecoacoustic tasks, including few-shot learning, is even better in newer general audio classifiers such as YAMNet (CNN) and AudioMAE (self-supervised transformer)

and best in embeddings from models pre-trained for global birdsong classification, such as BirdNET and Google Perch [174]. Loo et al. [177] also recently demonstrated that training a deep learning model to predict biologically-relevant information inherent to recording metadata – specifically, time of day – could encode ecological information in the model's embeddings, demonstrating prospects for supervised deep learning approaches that avoid manual labelling.

Collectively, these automated methods facilitate the acoustic monitoring of ecosystems from individual species to higher-level biodiversity and ecosystem health metrics across time and space at scale and pace, offering fertile ground for acoustically identifying tipping point EWSs. While the system state might not be monitored directly by such methods, they may encode it or provide suitable temporal and spatial proxies following local calibration and suitable pre-processing (see following subsection), just as satellite sensor products and indices, such as Vegetation Optical Depth VOD bands and the Normalised Difference Vegetation Index, provide proxies for biomass in vegetation resilience-sensing studies [62, 84, 139].

2.2.4 Step 4: Data Pre-Processing

Pre-processing is necessary to verify, clean-up and prepare acoustic monitoring data for calculating EWSs. Passive acoustic monitoring data will invariably contain data gaps from barriers to manual data collection (e.g., storms, harsh field conditions) or network outages and insufficient power for fully-autonomous deployments (e.g., solar panels receiving limited sunlight caused by extended cloud cover or obstruction from plant growth or snow). Therefore, down sampling or other methods of data aggregation/averaging, like sliding temporal windows, may be necessary, assuming this does not compromise the required temporal or spatial monitoring resolution. If resolution is compromised, data interpolation or imputation may be required, given data gaps can bias temporal EWSs including variance and autocorrelation [85, 180]. Interpolation or imputation methods should ideally preserve key statistical properties of the data, such as the overall or seasonal/diel mean, variance and autocorrelation, given these may influence or be directly used for the calculation of EWSs. Pre-processing to overcome imperfect data collection is commonplace for other sensing modalities, including satellite remote sensing, where calibration and other operations are necessary due to instrument degradation, cloud cover,

snow, shadows and other issues [85, 180]. However, while remote sensing can overcome data gaps by using multiple sensors [85], this is more challenging with acoustics' smaller spatiotemporal coverage, underscoring the benefits of monitoring replicate sites.

We also recommend inspecting acoustically-derived ecological metrics and raw audio to ensure that these are as anticipated for the study system's current stable state. For example: do acoustic indices capture expected spatial and temporal trends, such as peaks in avian activity during dawn and dusk choruses? Do deep learning embeddings follow similar distributions between site or season replicates? If patterns are unexpected, inspecting recordings may reveal underlying issues, such as poor signal-to-noise ratio from e.g., self-noise or microphone degradation [37]. Such issues might be rectified by applying appropriate frequency filters to remove undesired noise. It is also important to consider here what the appropriate signal is. Studies focussed on specific taxa may employ filtering and de-noising techniques that enhance the sounds from those species for more accurate monitoring [144]. In contrast, studies that are taxonomically broad or aim to detect ecosystem-wide resilience loss through soundscapes may need to incorporate a greater range of frequencies and level of background noise, particularly as the latter could be indicative of relevant features, such as insect activity [181] or acoustic reflections from vegetation [182] (which might be associated with the level of above ground biomass).

It is also common practice – and recommended [183,184] – to calibrate prediction score thresholds in species call classifiers to particular ecosystems and recording hardware by working with experts to manually label calls in a subset of recordings as a form of ground-truth [147, 172]. Thresholds that optimise performance (usually a trade-off between precision and recall – ensuring predictions are accurate and capture a large proportion of calls) can then be set [172] and/or species-specific probability thresholds derived [184].

Many temporal EWSs, including variance, autocorrelation and certain deep learning methods, require detrending of the system state (or state proxy) time series to separate long-term or seasonal trends from the short-term fluctuations in which critical slowing down occurs [185].[1] The fluctuations are usually obtained by taking the residuals of the detrending process. Detrending methods include simple subtraction of the long term

[1]It is worth noting that, conceptually, these fluctuations in ecological time series are effectively the noise around long-term trends, but in this pipeline, they represent the relevant signal in which to search for EWSs of critical slowing down. Accurately capturing these fluctuations may require appropriate pre-processing of the audio data to set a suitable signal-to-noise ratio for monitoring relevant system state variable(s) with acoustic indices, embeddings and species call classifiers, as discussed above.

mean [103, 186], Gaussian filtering [107, 186], and Locally weighted scatterplot smoothing (Lowess) [62, 187]. The nature of de-trending can impact EWSs [186]. For instance, Dablander and Bury [188] revealed optimal accuracy from a deep learning classifier for predicting bifurcations in time series data requires using the detrending method(s) applied to the model's training data. In the case of spatial EWSs, pre-processing can also be required, such as applying a discrete Fourier transform to spatial data [80] or to assemble data gathered along spatial gradients in a space-for-time substitution (e.g., [111, 112]).

2.2.5 Step 5: Searching for Early Warning Signals

The classic, most widespread temporal EWSs of critical slowing down are variance and autocorrelation [60]. These indicators are usually calculated on sliding windows that increment at a suitable resolution to capture system state fluctuations. The choice of window size can affect results, with longer windows providing smoother and more reliable EWSs [186]. However, it is important to note that for each increment of the window, the time series should be approximately stationary (i.e., the statistical properties, such as the mean, variance and autocorrelation do not vary over time). It is advisable to try a reasonable range of sliding window lengths to verify whether the calculated EWSs are robust to window size (e.g., [62]). Spurious trends in variance and autocorrelation unrelated to critical slowing down can arise. For example, variance is more susceptible to large perturbations or shocks and may be more likely than autocorrelation to increase due to 'one-off' events (e.g., a once-in-a-century drought) [62]. However, variance may also decrease near a bifurcation due to limited data or a reduction in external noise before tipping [189]. Autocorrelation is therefore generally thought to be a more robust indicator, though it can be influenced by changes to noise in the data. For instance, mixing newer and older satellite sensors for greater spatiotemporal coverage may increase autocorrelation, since uncorrelated measurement noise decreases as newer, less noisy sensors are introduced [190]. Therefore, it is recommended to ensure both variance and autocorrelation exhibit similar trends for robust results (e.g., [62]).

Other temporal EWSs include further statistical changes, such as spectral reddening (a shift in the power spectrum towards lower frequencies) [191] and an increase in skewness towards the system's prevailing alternative state [192]. Under strong stochastic forcing, the system may also 'flicker', i.e., make brief jumps to an alternate attractor [60]. Skew-

ness, variance and autocorrelation also have spatial equivalents [109,193] and complement other, more complex spatial EWSs such as Turing patterns (increasing self-organisation) and Kolmogorov complexity [60, 80], though it is less clear whether or how these could be captured acoustically. Further temporal critical slowing down-based methods include fitting statistical models, e.g., detrended fluctuation analysis [194], or using process-based models [60]. Deep learning models for time series classification also show potential. Bury et al. [187] developed a deep learning model that outperforms traditional EWSs at predicting bifurcations on a variety of model and empirical systems, while also predicting the normal form associated with the bifurcation (fold, transcritical or Hopf). This can offer vital insight into the nature of the transition and inform appropriate management strategies.

When monitoring data are temporally limited, other modelling and analytical approaches to inferring transitions and stability in dynamical systems may provide a viable alternative to traditional EWSs. For instance, when monitoring covers a brief period but is spatially extensive — ideally with multiple replicates of tipping (sub)systems (e.g., data from large annual ecological surveys) — it may be possible to construct hidden Markov models to assess the chance of tipping to an alternative state. Hidden Markov models provide a statistical means of inferring the hidden (i.e., unobservable) underlying states of a system over time, based on observable variables whose values depend probabilistically on the underlying state.[2] They assume that the underlying process follows the Markov assumption, that the current state depends only on the prior state(s)[3], and are widely used for modelling time series in many domains [195,196]. Vitense et al. [197] used hidden Markov models to estimate fold bifurcation transition probabilities from short monitoring time series in bistable lakes with clear and turbid alternative states (treated as the partially hidden states in their models). Their approach was also used to uncover the main drivers of such transitions. The models were built using annual nutrient, water quality and biomass data gathered across 123 lakes in Minnesota, USA, and represent an under-

[2]Note that here, 'state(s)' refers to the instance(s) of the hidden property underlying each point in a sequence or time series, which we wish to infer from the hidden Markov model. This 'state' is often discrete, and in the case of Vitense et al.'s study, it corresponds to a binary categorical classification of the alternate states of the system (i.e., clear or turbid). This classification categorises, but is different to, the actual continuous 'state' of dynamical system, which is a point that describes the value of a system's variable(s) at a given moment in time (and is how 'state' is usually used in this chapter).

[3]First-order models, which are typical, assume that the current state only depends on the previous state, or, put differently, that what happens next only depends on what is happening now. Higher-order hidden Markov Models assume dependence on multiple prior states.

explored but promising approach for understanding tipping risk in systems with known alternative states that currently lack long-term monitoring.

For multidimensional systems with stable oscillatory dynamics, Poincaré maps (a.k.a. return maps) may also provide a means to assess the loss of stability from temporally sparse observations. Poincaré maps specifically chart where the trajectory of an n-dimensional cyclic system state crosses an $(n-1)$-dimensional subspace in each cycle [69]. The stability of oscillations can be inferred from whether the successive intersection points converge or diverge. Hence, for systems that are at risk of tipping away from stable oscillatory dynamics (e.g., total collapse of a system with seasonal variation), observations of the system at the same point in each cycle (e.g., yearly acoustically-derived populations censuses), as opposed to continuous monitoring, could reveal the approach of such a transition.

Another highly promising class of EWSs that acoustics is well-poised to monitor are trait-based signals [105]. There is notable theoretical and empirical evidence that fitness-related, plastic phenotypic traits – namely, body size – may shift as environmental conditions change [105]. In microcosm studies, Clements and Ozgul [198] showed that the mean body size of a stressed population declined on a significant scale ahead of a collapse to extinction, while the population abundance displayed conventional EWSs including rising autocorrelation. Further work demonstrated similar trait and abundance signals in the past collapse of overharvested whale populations [199]. Acoustics may be able to detect such declines in body size in vocal animals through changes to the nature of vocalisations [200, 201]. Species call classifiers could, for instance, both identify the number of calls from different species, offering insight into relative abundance dynamics, and parse calls for analyses relating acoustic features (e.g., amplitude, fundamental frequency, nature of harmonics) to body size.

Overall, the development and observation of EWSs in natural systems has proliferated in the past two decades. We have highlighted certain prominent, acoustics-relevant examples, but broader reviews can be found in [80, 105, 202] and Dakos et al. [60], who also outline software packages for calculating EWSs. We expect that integrating passive acoustic monitoring into the study of EWSs may unleash another powerful route to understand and anticipate ecosystem tipping points.

2.2.6 Box 2: Applied Example

As a simple applied example for anticipating an ecological tipping point with acoustic data, consider the Asian songbird crisis (the trade of caged songbirds [203]). We might assume that the songbirds collectively undergo logistic population growth in the wild and their harvesting saturates with the captured songbird market and/or due to law enforcement restricting poaching, producing similar dynamics to Fig. 2.1b,d,f. If the songbird population growth and harvesting are balanced with the population at an upper abundance level (corresponding to equilibrium point x_+ in Fig. 2.1b,d), then gradually increasing the harvesting rate will erode the population's resilience and risks a sudden collapse to a much lower abundance (x_-, Fig. 2.1e). As this resilience is lost, fluctuations in the population abundance from natural minor variations or small perturbations (e.g., moderate droughts, wildfires, disease) will become larger and take longer to return to equilibrium (x_+), owing to critical slowing down ($\lambda \to 0$, Fig. 2.1d,g).

The number of songbird vocalisations is likely to reflect their population abundance and its slowing fluctuations, and could be inferred from long-term acoustic monitoring data in various ways. Vocalisations could be directly detected by a bird call classifier like BirdNET. While BirdNET may not pick up all vocalisations, if its performance is consistent (which should be achievable with suitable data pre-processing and species-specific calibration), then the subset of vocalisations it detects should exhibit slowing fluctuations that mirror those of the overall population as it undergoes critical slowing down.

Changes in songbird vocalisations will also alter the overall soundscape of Southeast Asian forests, which could be captured by suitable ecoacoustic soundscape descriptors, such as acoustic indices or machine learning model embeddings. The acoustic complexity and bioacoustic indices are in particular intended to capture the level of avian signing activity [152, 154], while embeddings from numerous models have been shown to predict avian richness [123] and encode various aspects of biodiversity [178, 204]. Both indices and embeddings could therefore show slowing fluctuations in songbird vocal activity, but must be applied and interpreted carefully. In particular, given that these descriptors can be influenced by other activity in soundscapes, pre-processing of the audio data (e.g., filtering) may be necessary to ensure the relevant signal (i.e., songbird vocalisations) is captured, and that changes to fluctuations in the descriptors are not driven by other vocal

species or differences in acoustic noise. Similarly, acoustic indices should be calculated on relevant time and frequency scales [144]. The nature of pre-processing and choice of analytical parameters will be specific to factors such as recording setup (type of microphone, sampling rate, compression etc.) and location (given the impact of vegetation on acoustic reflection and attenuation and influence of local sound-producing species on the soundscape). To verify whether the descriptors are capturing the relevant signal, their relationship to expected patterns should be inspected. For instance, in this case, the indices and embeddings would expected to peak at dawn and dusk during characteristic avian choruses at those times.

An advantage of acoustic indices is that they are arguably more tractable to interpret, given they can be more directly related to the spectral features of recordings from which they are calculated, than embeddings from 'black box' machine learning models. The cause of any unexpected patterns in acoustic indices could therefore be uncovered by manually inspecting spectrograms. However, embeddings' correlates could potentially indicate which soundscape features they correspond to. Further, embeddings have been shown to better predict aspects of ecosystem health and biodiversity than acoustic indices [123] and may therefore uncover more subtle, complex or multivariate patterns that 'human-derived' indices cannot. Embeddings are also more robust to certain recording parameters such as audio compression [176] and would thus be more suitable in instances where storing long-term uncompressed acoustic recordings is infeasible.

Lastly, to determine whether a tipping point in the songbird population is approaching, early warning signals can be calculated from the ecological time series generated by BirdNET and the soundscape descriptors outlined above. For early warning signals based on critical slowing down, the time series should generally first be de-trended to separate the short-term fluctuations from long-term trends. Classic statistical early warning signals such as variance and lag-1 autocorrelation could then be calculated, with increases in both indicating resilience loss towards a tipping point of abrupt songbird population decline. It may also be possible to apply a deep-learning bifurcation classifier to the time series [187], to determine the type of transition that could be approaching (in this example, a fold bifurcation would be expected). To strengthen findings and ensure that early warning signals genuinely reflect declining resilience, robustness testing of the analytical pipeline can also be conducted, e.g., trying different pre-processing techniques, varying

the time/frequency windows on which acoustic descriptors and early warning signals are calculated, and so on.

2.3 Opportunities and Challenges for Acoustically Monitoring Tipping Point Early Warning Signals

2.3.1 Acoustics May Reveal Early Warning Signals at Various Ecosystem Levels

Ecoacoustics can capture both detailed and high-level information about ecosystems, offering a myriad of routes to monitor salient ecological variables for EWSs. Which to prioritise may depend on the specific question at hand and/or outcomes of datamining large ecoacoustic datasets for EWSs. Species call classifiers can uncover the presence and activity of vocal species [14], and thereby reveal their relative population dynamics and the fluctuations in which critical slowing down manifests. Species abundances can also be directly estimated by using microphone arrays to determine the source direction or location of calls [42, 46, 205]. By distinguishing sounds arriving from different directions at a single point, autonomous recording units with small, integrated microphone arrays (e.g., [49]) could potentially even detect certain spatial EWSs or facilitate space-for-time substitutions with fewer recorders than sites – for instance by comparing variance and autocorrelation on ecoacoustic signals from neighbouring habitats.

There is a unique opportunity to monitor potential population collapses in terms of phenotypic trait changes that leave acoustic signatures in vocalisations [105,198–201]. Additionally, various soundscape analysis methods have to-date revealed different ecosystem responses to disturbance, including soundscape resistance and recovery [129–131,133,140]. Gottesman et al. [132] specifically note that soundscape variability could indicate landscape disturbance and, potentially, impending regime shifts.

Acoustic indices and deep learning model embeddings may also capture the state variable(s) of systems from communities to overall ecosystems, given their correlations with and ability to predict ecosystem health and biodiversity. While acoustic indices may be easier to relate mechanistically to the system state, embeddings' capacity to encode it could be uncovered by attempting to predict the system state variable(s), assuming

ground truth state data exists. A pertinent example could be predicting above-ground biomass, as model embeddings from tropical forest recordings have been used to do so before [123], and satellite-sensed vegetation data has revealed increasing EWSs of tipping points in global vegetation [84, 139], including tropical forests [62, 64, 113]. Whether the biomass predictions or EWSs derived from them match those from other modalities like satellite imagery could indicate how suitable certain model embeddings are for calculating EWSs.

As deep learning embeddings are high-dimensional, they could capture several variables in multivariate systems (e.g., multi-species population dynamics). The question for univariate or multivariate systems is in which embedding feature(s) to look for EWSs. One answer could be to use dimensionality-reduction techniques, which are not only used to apply univariate EWSs to multivariate systems [60, 103, 206], but also commonly applied to deep learning embeddings and other high-dimensional data for visualisation and lower-dimensional analyses (e.g., [207]). Principal component analysis, which seeks to capture much of the variance and overall directions of change in few new variables ('principal components' that linearly combine the original variables), may be particularly promising [60, 208]. Principal component analysis has been used in studies aiming to anticipate transitions in climate and biological systems [60, 103, 138, 206, 209] and could be applied to deep learning model embeddings in a parallel analysis to capture the overarching dynamics of a complex system. The first principal component could then be assessed for univariate EWSs and may suggest which embedding features contribute most to variability in the system based on the size of their loadings (coefficients). Those with the largest loadings could be investigated for EWSs and their correlations to biodiversity, to relate results back to the system's biology.

2.3.2 Acoustic Data are Limited but Growing

This represents both a challenge and an opportunity. On one hand, acoustic monitoring has to-date been of limited spatiotemporal scale compared to the large coverage needed to detect EWSs in many systems. Longer passive acoustic monitoring efforts are currently closer to years in scale than the decades needed to capture trends in forcing from human pressures and climate change. On the other hand, this is changing with the falling cost and increasing capability of tools for each stage of acoustic monitoring, from recorder

hardware to data storage and cloud computing (e.g., [41]). The advent of cheaper, robust autonomous recording units (e.g., AudioMoth [38, 210], Cornell's Swift/SwiftOne [39], BUGG (bugg.xyz) and its DIY predecessor [40]) has been instrumental in facilitating longer and larger deployments. Power efficiency gains (partly thanks to temporally subsetting recording) allow battery-powered units to now last for months before requiring manual battery replacement (e.g., [38]), while advances in fully-autonomous recorders, with self-sufficient powering and data transmission [40, 49], can unlock truly long-term, near real-time monitoring that minimises manual intervention post-deployment. Drone-deployed biodegradable sensors may increase the scale of future monitoring even further – especially in hard-to-access areas – while minimising environmental impacts [211]. In the marine realm, the recent discovery that whale vocalisations can be detected with fibre-optic telecommunication cables has potentially uncovered a global-scale, high resolution, real-time system for monitoring ocean health and biodiversity [51].

With the growing number of tools and datasets, however, comes the risk of disparate methods hindering collaboration and data assimilation for greater spatiotemporal coverage. Ecoacoustic metrics are generally local in nature and can be easily misinterpreted without study-specific calibration [37,159,184], while EWSs can be impacted by combining data from different sensors, such as inconsistent measurement noise biassing autocorrelation [190]. Further, while ecoacoustic monitoring tools are becoming less expensive, they remain costly especially for researchers based in less wealthy nations, where even costs associated with storing large passive acoustic monitoring datasets can be significant. To deal with this, raw data may need to be compressed or deleted after extracting relevant metrics, limiting future reuse of such datasets in large-scale collaborative monitoring studies.

These challenges can partly be alleviated by harmonising methods with the growing body of ecoacoustic monitoring guidelines. These include guides on compression and recording schedule [176], acoustic indices [37,149], species call classifiers [184], and overall monitoring pipelines [153,212]. Further, there are now various efforts to track global passive acoustic monitoring projects and assimilate their metadata, such as the 'Worldwide Soundscapes' project [116]. In addition, temporally-limited data might still be useful for identifying incipient tipping points if analysed with techniques like hidden Markov models [197] or Poincaré maps [69]. Patchy or small-scale recordings could also be com-

plemented by data from other monitoring modalities, such as remote sensing, camera traps, environmental DNA, LiDAR, and more [8]. There has in recent years been increasing work on multimodal environmental and biodiversity sensing (e.g., [213–215]) including pairing satellite imagery and high-throughput automated recorders [216]. With this, there is a growing need for multimodal EWSs that are more robust to data gaps, which could potentially be met by multimodal AI. There is also a notable opportunity to integrate multimodal monitoring data into more complex holistic ecosystem models, such as the nascent Virtual Ecosystem [217]. Such models could be used to simulate state transitions and/or provide high-resolution synthetic system state data on large spatiotemporal scales, with regular calibration from live monitoring, for calculating multitudinous EWSs.

2.3.3 Uncertainties in Tipping Point Theory and Early Warning Signals

Ecosystems and their resilience are complex [81,90,94,95,218,219]. Many ecosystems possess cyclic [81], transient [220], or chaotic dynamics [221, 222] that may not display typical critical slowing down-based EWSs or fold/transcritical bifurcations [134, 135] and the slow, broadly monotonic forcing they assume. Further, many understandings of ecosystem resilience might not conform to the tractable mathematics of lower-dimensional dynamical systems and bifurcation normal forms. Ecosystem resilience can encompass other ways of adapting to persist in function when faced with disturbance and changing conditions [73,94,223]. In such cases, EWSs not based on critical slowing down may be able to anticipate tipping points, but better models and quantitative metrics of the other mechanisms of ecological resilience are needed (e.g., [73,95]). Nonetheless, critical slowing down offers a strong basis for forewarning of tipping points for a remarkably wide range of real natural systems, including some with cyclic dynamics [81], hence it remains fundamental to assessing ecosystem resilience.

From a practical perspective, each step of monitoring EWSs may introduce potential uncertainties about the accuracy and reliability of results, from data gathering challenges (e.g., recorder self-noise, data gaps) that may hamper downstream analyses to the extraction of ecological metrics, data pre-processing, and EWSs calculation, all of which are sensitive to analysis parameters. Such uncertainties are particularly pronounced with the use of 'black box' deep learning tools, as the rationale for their predictions and 'meaning'

of their embeddings are not fully known (but can be explored with correlation analyses). The abstraction of these tools may call into question whether their combined outputs, such as applying a bifurcation classifier to a time series of the first principal component from acoustic deep learning embeddings, have any real meaning or biological basis.

However, many of these uncertainties can – and have – been investigated [186, 188] and overcome for other sensing modalities including remote sensing [62, 85, 190], leaving a clear roadmap for doing so with acoustic data. That includes grounding tipping point forecasting in a rigorous ecological understanding of the system and its capacity to tip alongside comprehensively robustness testing results and checking for consistency with EWSs from other sensing modalities. With such careful study design and validation, passive acoustic monitoring could uncover pertinent EWSs across scales.

2.4 Conclusion

As anthropogenic climate change and land use further intensify, the threat of numerous climate, ecological and Earth system tipping points is multiplying, with dire and potentially cascading consequences. To mitigate such consequences and even trigger positive tipping points that combat them – such as rapid rewilding and decarbonisation – we need tools for effectively anticipating tipping points and studying the health of affected systems more broadly. Acoustic monitoring has proven its capacity for the latter, and shows considerable promise for the former, given its ability to capture ecosystems at multiple levels, exceptionally high resolution and, in the case of fully autonomous recorders, in near real-time. While tipping point theory might not apply to all ecosystems and the scale and compatibility of ecoacoustic deployments are relatively small at present, there is now a crucial opportunity to integrate acoustics into tipping point forecasting systems, owing to the confluence of: (i) increasingly capable, low-cost hardware powering larger acoustic monitoring deployments; (ii) advancing tools and guidelines for the automated extraction of ecological information from soundscape recordings; and, (iii) improving confidence and methods in tipping point EWSs. Leveraging acoustics to this end may prove vital in detecting the collapse of ecosystems that underpin nature and society just in time to intervene.

Chapter 3

Towards using virtual acoustics for evaluating spatial ecoacoustic technologies

Summary

1. Small microphone arrays and sound-source localisation algorithms are increasingly prevalent in the passive acoustic monitoring of ecosystems. These technologies enable analysis of natural soundscapes' spatial features, yielding additional insights into biodiversity and ecosystem health. While many of these technologies have been evaluated in the field, there is a lack of controlled, repeatable methods to test them.

2. We developed an ambisonic Virtual Sound Environment (VSE) for simulating real natural soundscapes to evaluate spatial passive acoustic monitoring technologies. We validated this novel approach using a passive acoustic monitoring recorder with a six-microphone array, from which we extracted a typical suite of ecoacoustic metrics, including acoustic indices and avian species predictions and localisations from the software BirdNET and HARKBird, respectively. We first verified whether the VSE could replicate natural soundscapes well enough to test passive acoustic monitoring technologies by comparing these metrics between field and VSE-based recordings. To pilot the VSE as an environment for testing

passive acoustic monitoring hardware, we assessed how orientation impacts the six-microphone array's performance by using the same suite of metrics to compare VSE recordings made with the array at various pitch angles. Finally, we piloted the VSE as a test platform for passive acoustic monitoring software by investigating how BirdNET and HARKBird perform on bird calls added to the VSE-replicated soundscapes.

3. While the VSE and field recordings had similarities in some metrics, including spectral composition and BirdNET predictions, ambisonics' perceptual bias and susceptibility to spatial aliasing limited the spatial analyses that could be undertaken. Our trials nonetheless revealed that device orientation impacts the performance of HARKBird and certain ecoacoustic indices, and that BirdNET and HARKBird perform best on louder, more directional bird calls.

4. Our results demonstrate the potential for this approach, but highlight limitations to using an ambisonics-based VSE. We thus provide guidelines for the use and refinement of such systems towards more standardised, controlled benchmarking of passive acoustic monitoring technologies, empowering practitioners to make more informed decisions on using these vital tools.

3.1 Introduction

Spatial audio technologies are increasingly widespread for the passive acoustic monitoring of natural environments. Regarding hardware, this primarily consists of multi-microphone recording approaches. These encompass both distributed microphone arrays, which consist of several typically-independent microphones spread across space [44, 46, 224, 225], and, increasingly, integrated microphone arrays, in which two microphones [226, 227] or more (e.g., [53, 228, 229]) are contained within a single, multichannel recording device. Regarding software, sound-source localisation algorithms continue to improve, partly thanks to deep learning methods [50] and the development of localisation applications for animal vocalisations, such as the avian call localisation tool HARKBird [53, 54]. Together these spatial passive acoustic monitoring hardware and software developments are providing

superior metrics of ecosystem health and biodiversity [229], including species abundance estimates [46, 205].

While many of these technologies have been tested in the field, real natural soundscapes are inherently and continually variable, and somewhat unpredictable, limiting the control and replicability of these test methods. This can partly be avoided with playback experiments, in which one or more loudspeakers play back pre-recorded sounds in the field [25, 58, 230]. Further control and replicability can be achieved by simulating more of the field recording process, but few methods for this have been proposed – especially for testing spatial passive acoustic monitoring technologies. Kaneko and Gamper [231] recently developed an approach to simulate distributed microphone array field recordings of a large forest and test their robustness to noise, reverberation and measurement errors, such as imperfect microphone placement. However, as their approach involves simulating field recordings, it cannot be used to test spatial passive acoustic monitoring hardware. Moreover, their approach is not intended for testing multichannel recorders with integrated microphone arrays.

Other fields, such as audiology, have leveraged various sound field reconstruction techniques to create Virtual Sound Environments (VSEs) [232–234] that replicate real spatial sound-scenes in laboratory settings, typically to evaluate human hearing and test multi-microphone hearing aids. VSEs have been employed in audiology to overcome the limited control and reproducibility of real-world tests [56, 57] – challenges which also arise for field testing passive acoustic monitoring technologies. Several studies have used ambisonics, a technique for creating, recording, modifying and reproducing spatial sound fields in an efficient and flexible manner, to create VSEs for evaluating hearing aids [57, 232, 234, 235] and measuring certain features of speech intelligibility and perception [233, 236].

In ambisonics, sound fields are decomposed into spherical harmonic components (akin to a Fourier series for 3D space [237]), starting with an omnidirectional component, plus a number of directional components, each corresponding to a specific directional pattern, for example bi-directional (front-back, left-right, up-down) for the first order, and more complex ones for higher ambisonic orders. Monophonic signals can also be encoded to the spherical harmonic domain, allowing additional sources to be added to any approximate location within a sound field. This spherical harmonic representation can then be decoded to any suitable loudspeaker array, provided a sufficient number of regularly-spaced

loudspeakers ($(n+1)^2$, where n is the ambisonic order) [238–240]. First order ambisonics was primarily developed by Michael Gerzon [241, 242]. It is often used today for entertainment, such as rendering spatial audio over headphones [240], but has a small 'sweet spot' in which the sound field is most accurately reproduced [57, 234, 243]. Higher-order ambisonics has a larger sweet spot, but requires more microphones, loudspeakers and processing for recording and reproduction [239, 240].

Even for higher orders, the sweet spot's size is limited, in part due to spatial aliasing during recording and playback [57, 239]. This amounts to distortions between sound waves coming from different directions [244], which can affect their expected arrival time and level differences within the reconstructed sound field [234], leading to inaccurate directionality. Higher-order ambisonic playback can suffer from aliasing as it always uses multiple loudspeakers to reproduce signals, even for single point sources [57]. Furthermore, ambisonics was primarily designed around the psychoacoustic properties of spatial sound rather than producing a precise physical replica of spatial sound fields [241, 245]. As such, ambisonics reconstructs sound fields in a perceptually accurate manner for human listeners [243, 245], but may be physically imprecise, even in the sweet spot. This underpins some of the above inaccuracies in ambisonic reconstruction as well as other spectral, spatial and temporal artifacts it can create [232].

Nonetheless, ambisonics is a powerful tool for capturing, altering and reproducing real, complex spatial sound fields (including their reverberation and other acoustic properties) instead of the more labour-intensive method of reconstructing them with room acoustics models, for example [232]. This has driven ambisonics' uptake for applications requiring VSEs to recreate dynamic, noisy environments, and should be suited to reproducing natural soundscapes, whose acoustics are challenging to model.

We therefore developed a higher- (specifically, third-) order ambisonic VSE, consisting of a spherical twenty-five loudspeaker array, for reproducing real spatial natural soundscapes to evaluate spatial passive acoustic monitoring recorders and software in a controlled, replicable manner. There has been some work on multi-loudspeaker playback in ecoacoustics, namely for studying avian behaviour [225, 228, 246]. However, there is little work involving full-sphere sound field recording and reconstruction in ecoacoustics, beyond more artistic and experiential applications [247]. Moreover, we demonstrate that full-sphere reconstruction can be achieved in a comparatively resource-efficient way,

through repurposed or increasingly accessible equipment.

Here, we detail this VSE test platform and work to validate it (Fig. 3.1). This followed three sub-projects, to investigate whether the VSE could:

1. **Soundscape Reproduction Efficacy:** Effectively simulate real, spatial natural soundscapes for testing passive acoustic monitoring technologies. We verified this by comparing a standard suite of ecoacoustic metrics between recordings of real and VSE-replicated soundscapes.

2. **Impact of Device Orientation:** Test passive acoustic monitoring hardware and factors important to spatial passive acoustic monitoring deployments. As a proof-of-concept for using the VSE as a test environment, we re-recorded VSE soundscapes with MAARU (Multichannel Acoustic Autonomous Recording Unit; [49]), a six-microphone passive acoustic monitoring recorder, oriented at various pitch angles, to study how device orientation impacts the same standard ecoacoustic metrics. These include the popular avian classification algorithm BirdNET [167] and the avian localisation tool HARKBird [54]. Together, these applications can provide lower-bound avian species abundance estimates [205] by revealing whether there are several birds of a particular species calling simultaneously from different locations around a microphone array. However, HARKBird is expected to be susceptible to device orientation. Hence, we explore whether the VSE could reveal the degree to which HARKBird localisations are altered by MAARU's orientation.

3. **BirdNET and HARKBird Performance on Additional Avian Calls:** Test passive acoustic monitoring analysis software. To demonstrate this, we used the platform to explore the performance of BirdNET and HARKBird on additional avian calls embedded in the VSE soundscapes. We first added calls by ambisonic encoding; however, given the limitations of ambisonics' spatial accuracy, we also reproduced calls' direct paths from individual loudspeakers within the VSE and compared both embedding methods.

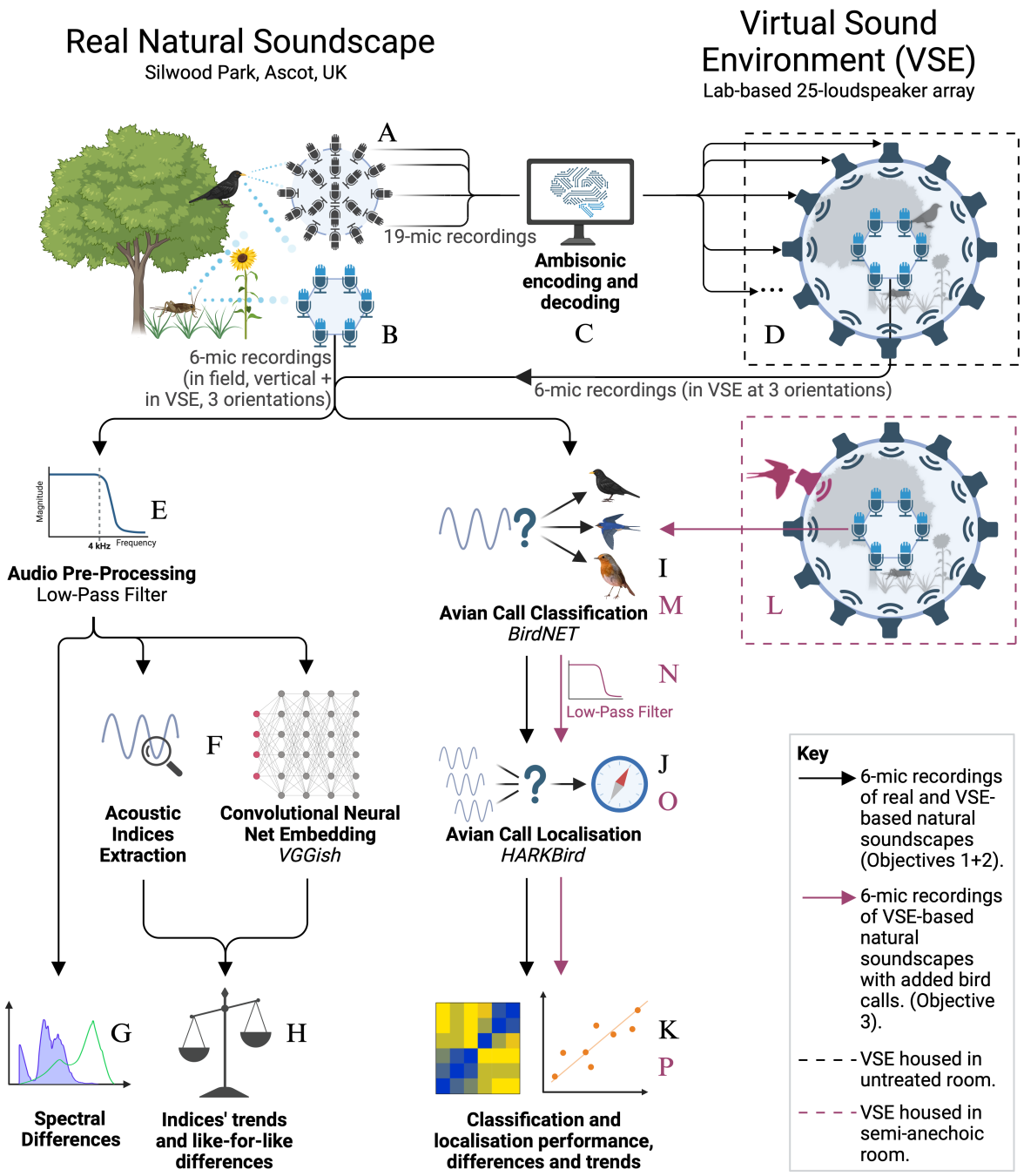


Figure 3.1: System overview and experimental structure. We captured real, spatial natural soundscapes with a high-quality 19-microphone array (**A**) and MAARU, the six-microphone passive acoustic monitoring recorder being tested (**B**). The 19-mic recordings were encoded to the ambisonic domain and decoded to a 25-loudspeaker VSE (**C**), allowing the soundscapes to be re-recorded by MAARU (**D**) with adjustments to relevant recording parameters, namely, device orientation. The six-microphone recordings from the field (**B**) and VSE (**D**) were passed through a suite of typical ecoacoustic metrics (**F, I, J**). Various analyses (**G, H, K**) were performed to compare the field and VSE recordings and explore the impact of device orientation. The process was repeated for VSE recordings with additional bird calls (**L**), to further explore the performance of software tools BirdNET and HARKBird (**M**–**P**).

3.2 Materials and Methods

3.2.1 Platform Description

The concept of this test platform is to take high quality, spatial recordings of natural soundscapes and use a spatial sound field reconstruction method to accurately reproduce them in a controlled environment (Fig. 3.1A,C). One can then use the passive acoustic monitoring hardware and software being tested to repeatedly record and analyse these 'virtual' soundscapes under various conditions (Fig. 3.1D,L). For example, one can alter the set-up of the passive acoustic monitoring recorder or modify the reconstructed soundscape to simulate effects like the presence of additional species (Fig. 3.1L). This also enables passive acoustic monitoring technologies to be tested on compatible spatial recordings of soundscapes from diverse and challenging environments (e.g., hard to reach or stormy areas), all within a single lab setting.

We adopted a flexible, resource-efficient approach to implementing the above concept. We used the 19-channel *ZYLIA* ZM-1 microphone array [248] with *ZYLIA*'s recording software (version 2.1.2, [249]) to take high quality recordings of various natural soundscapes (Figs. 3.1A, 3.2). The ZM-1 is considerably more affordable than other higher-order ambisonic recorders such as *MH Acoustics*' 'Eigenmike' devices.[1,2] We used *ZYLIA*'s software to convert the recordings to third-order ambisonics (SN3D normalisation, Furse-Malham channel ordering) (version 1.6.1; [250]) and reproduced these ambisonic recordings through an array of 25 pre-owned *Genelec* 8010A loudspeakers[3] (Fig. 3.1C), mounted to a sphere constructed from two commercially-available, steel hemispherical frames (Fig. 3.2).[4] The third-order ZM-1 recordings were decoded to the loudspeaker array using the free ICST Ambisonics externals for *Max/MSP* [251].

For most of this work the VSE was housed in a non-acoustically-treated room (Figs. 3.1 and 3.2), mirroring conditions in many ecoacoustics labs, given their focus on field recording and high cost of creating anechoic/soundproof spaces. However, this room likely introduced artifacts from reverberation, room modes and so on.

[1] https://mhacoustics.com/products

[2] https://www.creativefieldrecording.com/2017/03/30/ambisonic-field-recording-resources/#:~:text=MH%20Acoustics%20EigenMike%20em32%20(%24,via%20the%20EigenMike%20Interface%20Box

[3] https://www.genelec.com/8010a

[4] https://www.lifetime.com/lifetime-90136-dome-climber-earthtone

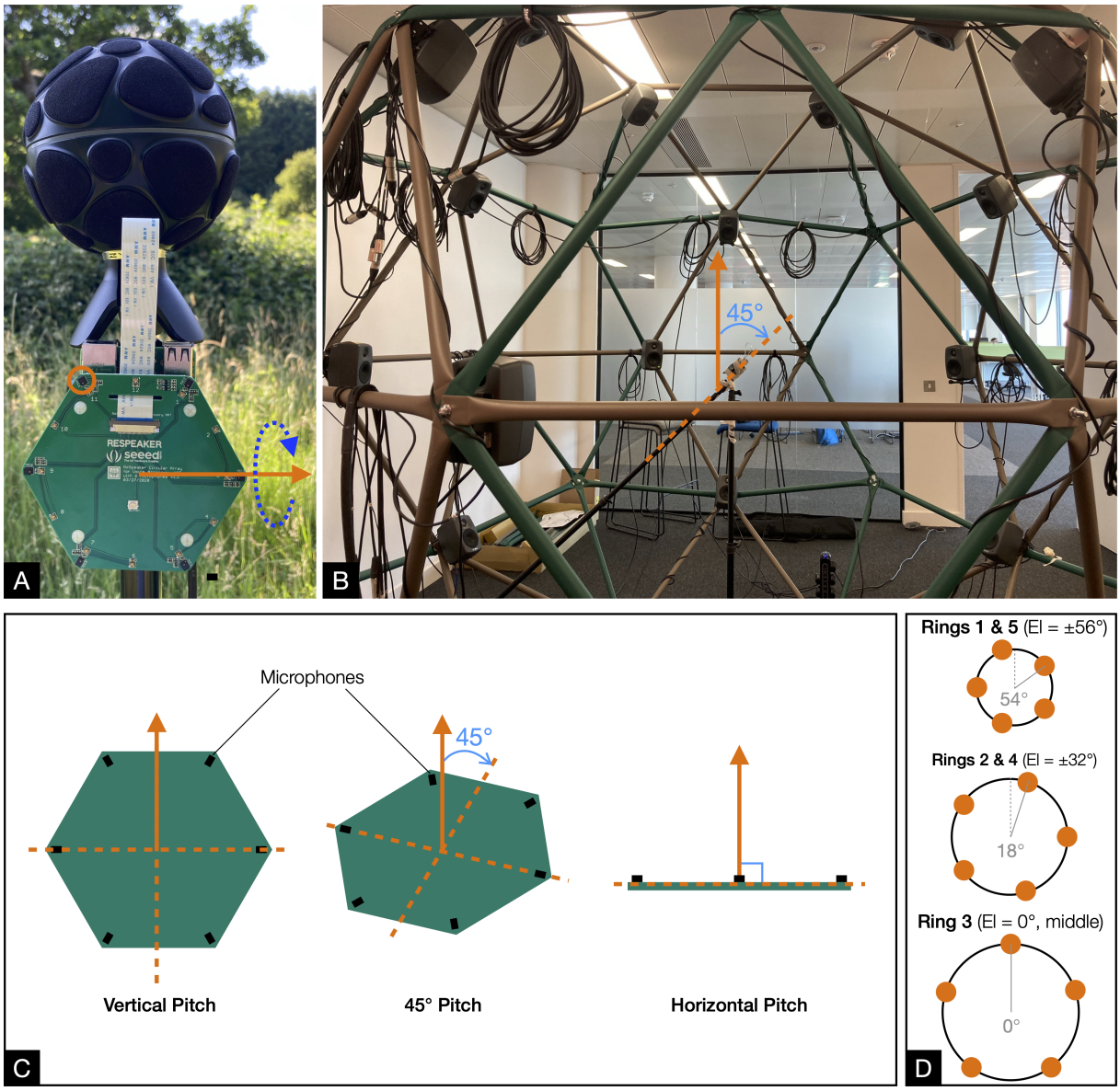


Figure 3.2: **A.** Field recording setup with MAARU, the six-microphone passive acoustic monitoring recorder being tested, beneath the spherical 19-mic ZM-1 used for ambisonic recording. Recordings for all single-channel analyses were obtained from Mic 1 of MAARU, circled. The arrows show the rotation axis and direction to change MAARU's pitch angle in the VSE recordings for Sub-project 2. **B.** VSE in the lab space it was first housed in, with MAARU oriented at 45° pitch. **C.** MAARU's pitch angles for the VSE recordings. **D.** Arrangement of loudspeakers on the VSE. The VSE consists of five rings at the elevations indicated (El - Elevation). Each ring contains five evenly-spaced loudspeakers, the first of which is at the azimuth indicated inside the rings.

Calibration

The VSE was calibrated in two steps to reproduce the same approximate Sound Pressure Level (SPL) as the original soundscapes. First, the 8010A loudspeakers were calibrated relative to one another. A calibrated SPL meter was placed in the centre of the VSE and

each loudspeaker's gain was adjusted so it reproduced a 1 kHz sine wave to around 75±1 dB. Second, the entire array was calibrated relative to the ZM-1's built-in gain. White noise was encoded to third-order ambisonics at zero amplitude and elevation, decoded through the VSE and monitored with the SPL meter. It was then recorded with the ZM-1 positioned in the sphere's centre, and aligned at 0° azimuth and elevation. Next, the ZM-1-recorded white noise was reproduced through the VSE. Using the SPL meter, the gains of all loudspeakers were adjusted so that the ZM-1's recording of the white noise was reproduced within around ±1 dB of the original white noise. The required adjustment was around −12 dB (which means, prior to adjustment, sound sources recorded with the ZM-1 would appear around four times closer than they actually were, assuming free-field spherical sound propagation). This second step was repeated for a 1 kHz sine wave, which also yielded an approximately −12 dB adjustment for all loudspeakers.

3.2.2 Recording Devices, Locations and Schedule

We piloted the VSE on MAARU [49], a six-microphone spatial passive acoustic monitoring recorder (Fig. 3.1B, D), with its weatherproofing enclosure removed. MAARU advances Sethi et al.'s open-source, monophonic passive acoustic monitoring device [40] with *Seeed Studio*'s affordable 'ReSpeaker 6-Mic Circular Array Kit' [252] which features six omnidirectional MEMS microphones on the vertices of a flat hexagonal PCB. To restrict file size while still allowing for a breadth of analyses, we used MAARU's intended 16 kHz sampling rate.

We conducted field recordings in July 2022 at Imperial College London's Silwood Park campus (Ascot, Berks., UK), using MAARU oriented vertically and the ZM-1 (sampling rate 48 kHz) mounted directly above it (Figs. 3.1A, B and 3.2). We took 10-min recordings at each of 6 sites spanning various levels of forest density and proximity to a stream or human activity. Government-issued fieldwork licences were not required for this work.

We re-recorded the simulated soundscapes using MAARU at the centre of the VSE (Fig. 3.1D), initially oriented vertically (as in the field), then at 45° pitch, and finally horizontally (90° pitch; Fig. 3.2).

3.2.3 Soundscape Reproduction Efficacy and Impact of Device Orientation

We extracted a standard suite of ecoacoustic metrics from MAARU's field and VSE recordings (240 mins total, Table B.1) to compare MAARU's recordings towards Sub-projects 1 and 2 (Fig. 3.1F–K). This comprised spectral power differences, seven acoustic indices, feature embeddings from an audio-specific CNN, and avian species predictions and localisations from BirdNET and HARKBird, respectively. Apart from HARKBird, these metrics use monophonic recordings. Such metrics are generally consistent across MAARU's microphones due to MEMS microphones' reliability and their proximity on MAARU's array. Indeed, the median variance of signal RMS values across MAARU's six microphones for all recordings in the 240 mins we captured was 0.366 dB (range: 0.062 to 8.892 dB). We thus extracted all monophonic metrics from MAARU's upper left microphone.

Spectral Differences, Acoustic Indices and VGGish Features

We calculated power spectrograms for all the recordings with 'pspectrum' in *MATLAB* (version 2022b), using 0.5 s Kaiser windows with 10% overlap, 0.7 leakage and 1024 Discrete Fourier Transform points. We then subtracted the field recordings' power spectrograms from those for the VSE recordings. In around two thirds of all recordings, we initially noticed a substantial increase in power around 7-8 kHz. This was likely introduced by background noise in the non-anechoic lab environment that may have been exacerbated by the frequency responses of the ZM-1 and *Genelec* loudspeakers. To remove this, we low-passed all recordings with a 4 kHz cutoff and 12 dB roll-off (Fig. 3.1E). The low-passed data were used for most subsequent analyses except for BirdNET and HARKBird's use for Sub-project 2, as most real bird calls in these recordings were too attenuated by low-passing to be detected. This, however, was different for HARKBird's use for Sub-project 3 (Fig. 3.1O, P) where low-passing was beneficial (see Section 3.2.4).

We computed the following canonical acoustic indices [176, 253, 254] on continuous 30 s windows[5] using the *R* packages *seewave* (version 2.2.0; [255]) and *soundecology* (version 1.3.3; [256]): Acoustic Complexity Index (ACI), Acoustic Diversity Index (ADI),

[5] 60 s or more is a more common window size [254], but would lead to too few data points given the brevity of our recordings.

Acoustic Evenness (AEve), Bioacoustic Index (Bio), Acoustic Entropy (H), Median of the Amplitude Envelope (M), and Normalized Difference Soundscape Index (NDSI). We also used the 'vggishPreprocess' and 'predict' functions from *MATLAB*'s *Deep Learning Toolbox* (version 14.5; [257]) with default parameters to compute the 128 dimension embedding of VGGish, a CNN pre-trained on a broad corpus of labelled audio samples from YouTube [163]. This embedding has been highly effective for numerous ecological analyses and classification tasks [123, 176].

We performed an adapted Bland-Altman analysis [258] to examine scaled differences in acoustic indices and VGGish features between field and VSE recordings. Bland-Altman analysis assesses the agreement between two methods of measurement and involves evaluating the difference between both measures to identify bias. This bias can be measured relative to the average of both measures, or one of the measures if it can be considered a reference. 'Limits of agreement' defining upper and lower bounds for acceptable bias can then be set based on context or differences' 95% confidence interval if they are normally distributed [258].

We used the field recordings as reference. However, the indices/features studied vary considerably in magnitude and range, lack clear contextual limits of agreement and were not normally distributed. Therefore, per [176] and [259], we calculated indices/features' differences as a percentage of their range in the field recordings. We zeroed differences for VGGish features with zero range in field recordings to avoid singularities, and averaged the VGGish differences across all 128 features to obtain a univariate difference set akin to other indices. As an ideal VSE should show minimal differences, we followed [176] in using ±5% limits of agreement and considering data strongly altered if their interquartile range omitted zero. We also explored ±2.5%, ±7.5% and ±10% limits of agreement, as a ±5% difference is challenging to interpret ecologically, given the extracted indices/features' correlations with biodiversity are highly context-dependent and may vary by recording/playback equipment [159]. Further, while differences were expected to be consistent across VSE recordings, we also assessed for correlations between MAARU's VSE pitch angle and indices' scaled differences.

BirdNET and HARKBird

We ran unfiltered recordings from MAARU's upper left microphone through BirdNET and from its six microphones through HARKBird (Fig. 3.1D). We passed the coordinates of each field site to the BirdNET graphical user interface to restrict it to birds from this region, and used default values for the other parameters. While BirdNET results would be similar across recording conditions with an ideal VSE, BirdNET's confidence score – which often positively correlates with prediction accuracy – can vary greatly by species and recording/playback hardware [184]. We thus restricted all BirdNET predictions to those with a greater than or equal confidence to the average for their species in the corresponding field recording. Species only predicted in the re-recordings were filtered using the corresponding site's overall average field recording confidence. After filtering, recordings had 20 predictions on average (range = 0-54). We filtered the HARKBird localisations to those whose timestamps overlapped with the filtered BirdNET predictions, and then found the percentage 'overlap' by calculating the proportion of field recordings' predictions and localisations that appeared in the VSE re-recordings. We used inferential statistics to explore differences in BirdNET predictions' confidence values and HARKBird's localisations between all recording conditions. We also evaluated correlations between device orientation and confidence values and localisations.

To compare BirdNET's precision and recall across all recording conditions, a professional ornithologist, well-experienced with conducting UK bird surveys, manually labelled the omnidirectional channel of the ZM-1's field recordings using non-overlapping 3 s windows to match BirdNET. As desired precision and recall vary by application, we calculated them for a range of confidence thresholds: 0.25 (average predictions per recording = 14, range = 0-35), 0.5 (average predictions per recording = 9, range = 0-29) and 0.75 (average predictions per recording = 5, range = 0-20; Fig. 3.6), and 0.01 and 0.9 (Table B.2).

3.2.4 BirdNET and HARKBird Performance on Additional Avian Calls

We performed 10 further VSE re-recordings (100 mins total, Table B.1), with MAARU oriented horizontally and additional avian calls embedded at certain spatial locations in the first five sites' ambisonic field recordings (Fig. 3.1 E, Table 3.4) during quieter

passages. We then verified how accurately these were identified by BirdNET and localised by HARKBird.

We added six common English bird vocalisations (Table 3.4) from audio recordings freely available on Xeno-Canto [260–265]. We first added the calls' direct paths to the soundscapes by encoding them to the ambisonic domain using the ICST Ambisonic externals in *Max/MSP* [251], enabling them to originate from any direction. To avoid spatial aliasing in ambisonic reconstruction - which can be particularly problematic for HARKBird's underlying localisation algorithm, MUSIC [53, 55] - we also reproduced the direct paths from individual loudspeakers closest to each ambisonic encoding position, providing superior directionality.

In both cases, we added a simulation of each field site's reverberation, following a method outlined by [266]. The reverberation is produced by convolving the direct path of each added sound with the zero-padded Impulse Response of the acoustic reflections from the corresponding site. We found the impulse response at a distance of 2m at 0°, 90°, 180° and 270° about the ZM-1 microphone for all but the final site, which had insufficient space about the microphone to do so. Each added bird vocalisation was convolved with the impulse response measured closest to its azimuth.

We initially noticed inaccurate HARKBird localisations for both encoding methods, suggesting spatial aliasing between MAARU's microphones due to its size and geometry [244]. MAARU's closest microphones are approximately 4.5 cm apart, hence aliasing could occur for sounds with a wavelength of 9 cm or less – roughly 3800 Hz and above. We therefore reused the 4 kHz low-pass filter from previous analyses on these recordings before passing to HARKBird (Fig. 3.1N)[6], notably improving localisation accuracy. To verify other factors did not cause the inaccuracies, we ran HARKBird on unfiltered recordings of bird calls added from individual loudspeakers without reverberation, and with neither reverberation nor the background ambisonic soundscapes (Fig. B.2). However, performance was best on low-passed, single loudspeaker recordings (with reverberation and the background soundscape), so we ultimately compared this approach to the ambisonic encoding.

After passing the recordings through BirdNET and HARKBird, we investigated the

[6]We did not apply the 4 kHz low-pass filter to recordings passed to BirdNET and HARKBird for Subprojects 1 and 2, as the real bird calls in these recordings were too quiet for detection after low-passing, even with a 6-12 dB gain boost (beyond which clipping would occur).

proportion of added calls that were accurately classified and localised and how the prediction confidence and localisation error varied by calls' elevation and embedding method.

3.3 Results

3.3.1 Soundscape Reproduction Efficacy and Impact of Device Orientation

Spectral Differences

There were progressively greater overall level increases for the Vertical, 45° and Horizontal re-recordings (Table 3.1), particularly in the 0-5 kHz bands. These increases were generally up to ∼12-13 dB for lower frequencies, meaning individual sound sources within these frequencies may have appeared up to around four times closer to MAARU in the VSE than they were in the field (assuming free-field, spherical sound propagation). However, the average variance in spectral difference across all frequency bands for all re-recordings was only around 1.6 dB (range 1.49 to 1.71 dB). To exemplify these patterns, data from Sites 2 and 5 are shown in Fig 3.3.

Table 3.1: Average differences in spectral power (dB) per 1 kHz band (up to the 8 kHz Nyquist frequency) between all VSE re-recordings and the reference field recordings. To calculate the mean and variance values presented, we first averaged the spectral differences for every site into 8 bands by taking the mean along the frequency axis for each 1 kHz segment (equivalent to 128 rows per segment, since 1024 discrete Fourier transform points were used to compute the underlying spectrograms). We then computed the temporal mean and variance of each 1 kHz band, and present the mean of these values across all 6 sites below.

	Vertical		**45°**		**Horizontal**	
Frequency Band	*Mean (dB)*	*Variance (dB)*	*Mean (dB)*	*Variance (dB)*	*Mean (dB)*	*Variance (dB)*
0-1 kHz	12.19	2.37	12.78	2.30	12.96	2.28
1-2 kHz	9.02	0.76	11.45	0.53	13.09	0.58
2-3 kHz	5.54	1.82	10.67	1.10	13.54	1.23
3-4 kHz	4.49	2.40	7.62	1.64	12.41	1.73
4-5 kHz	4.70	1.86	6.35	1.95	12.32	1.86
5-6 kHz	1.67	2.61	3.34	2.30	7.85	1.86
6-7 kHz	0.73	1.26	1.31	1.35	2.61	1.68
7-8 kHz	0.05	0.63	0.06	0.78	0.09	0.97
Average across all frequency bands	4.80	1.71	6.70	1.49	9.36	1.52

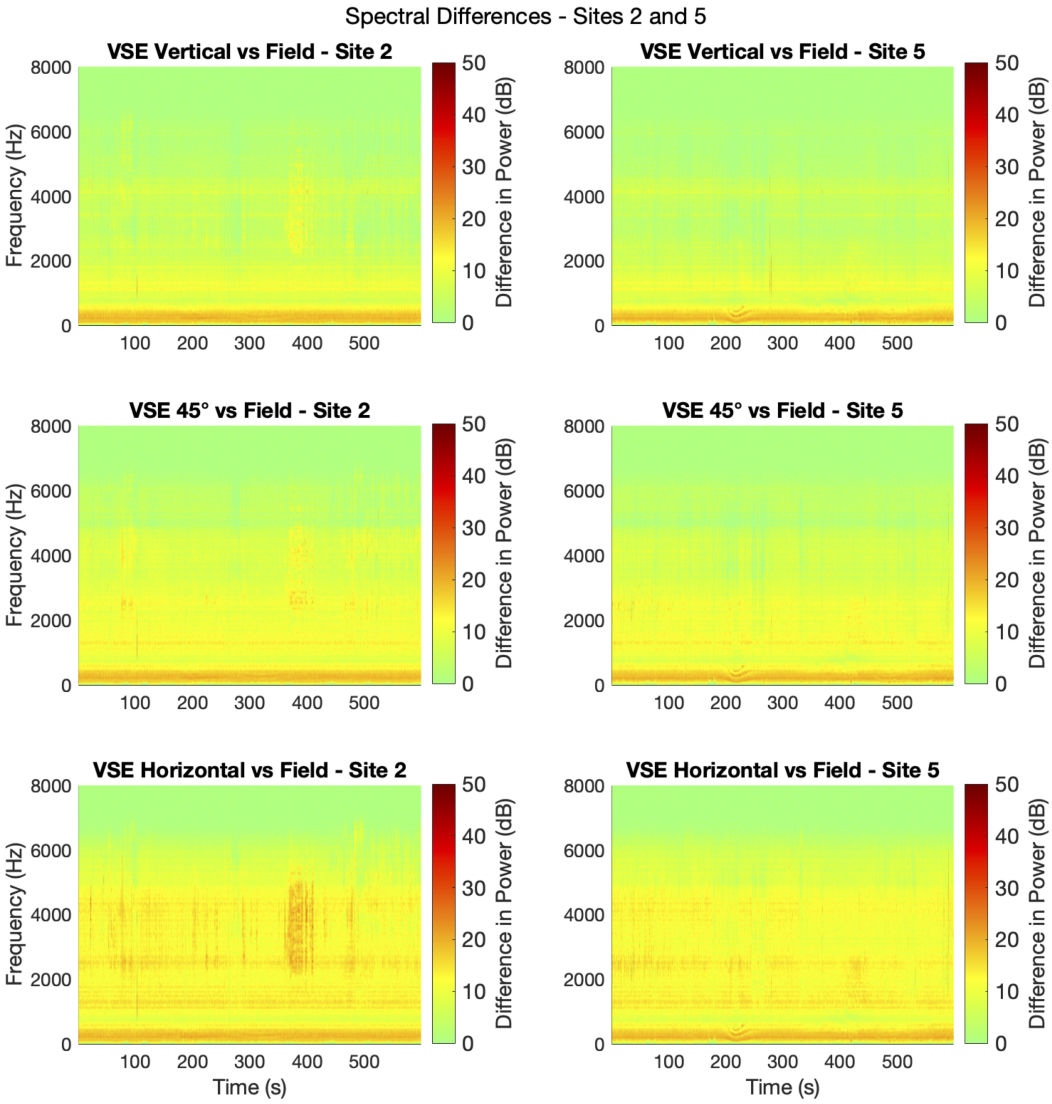


Figure 3.3: Differences in spectral power between each VSE re-recording and the field recording (audio captured from top left microphone of MAARU and low-passed above 4 kHz with a 12 dB roll-off) for Sites 2 and 5, which exemplify patterns seen across all sites.

Acoustic Indices and VGGish Features

In our adapted Bland-Altman analysis, only the VGGish features remained within ±5% difference to the field recordings across all VSE recordings, while the acoustic indices varied considerably (Figs. 3.4, B.1). These trends were similar for other limits of agreement, with only ACI and M appearing acceptably altered under the most generous alternative limits. Interestingly, absolute percentage differences rarely increased as re-recording orientation progressively changed. Rather, most decreased or showed no trend. Significant Spearman's rank correlations were found for all indices but M, the strongest being for ADI, AEve, NDSI and H (Table 3.2).

BirdNET and HARKBird

Device orientation impacted the performance of BirdNET and HARKBird. For BirdNET, VSE re-recordings overlapped with the field recordings by 51.2% (vertical), 40.4% (45°) and 55.4% (horizontal) on average. Fig. 3.5 shows spectrograms of vocalisations that were repeatedly correctly predicted by BirdNET in each recording condition, but the predictions did not always occur at the same times (i.e., different overlap). For HARKBird, allowing ±45° in localisation error, average overlaps were just 2.8% (vertical), 4.7% (45°)

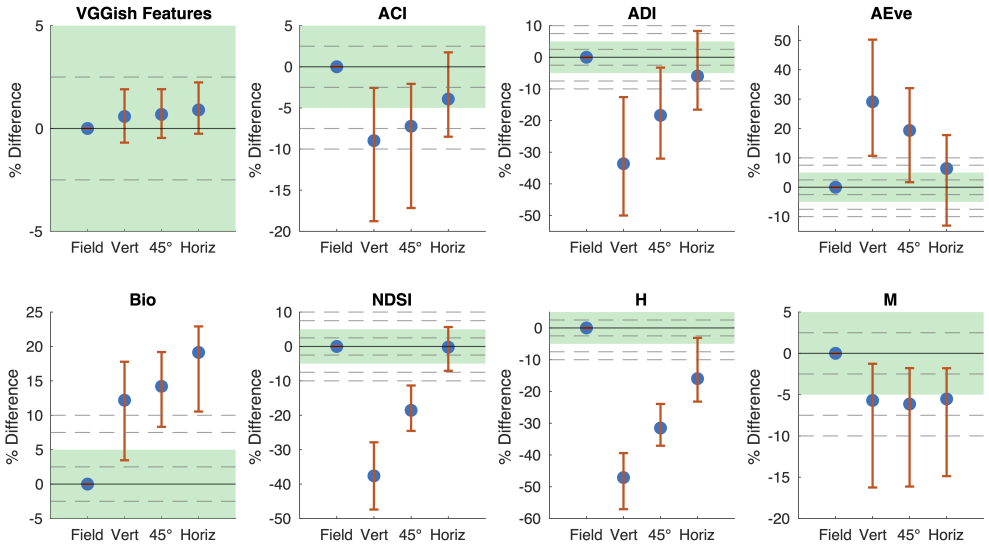


Figure 3.4: Plots of indices' average scaled like-for-like differences across all sites. Blue dots indicate the median percentage differences, whiskers show the upper and lower quartiles, and the shaded area denotes the ±5% limits of agreement. Where appropriate, the grey dashed lines represent ±2.5%, ±7.5% and ±10% as alternative limits of agreement.

Table 3.2: Spearman's rank correlation coefficients and p-values assessing the relationship between MAARU's orientation and: (i) acoustic indices' and VGGish features' scaled like-for-like differences; (ii) HARKBird localisations; and, (iii) BirdNET confidence values (for all detected birds and each of the four most predicted species).

	Rho	***p*-value**
Mean VGGish Features	**0.077**	**< 0.001**
ACI (Acoustic Complexity Index)	**0.238**	**< 0.001**
ADI (Acoustic Diversity Index)	**0.428**	**< 0.001**
AEve (Acoustic Evenness)	**−0.392**	**< 0.001**
Bio (Bioacoustic Index)	**0.238**	**< 0.001**
NDSI (Normalised Difference Soundscape Index)	**0.772**	**< 0.001**
H (Acoustic Entropy)	**0.568**	**< 0.001**
M (Median of the Amplitude Envelope)	−0.013	0.812
HARKBird localisations	**−0.583**	**0.001**
BirdNET confidence - all *(19 species, 205 predictions)*	−0.005	0.944
BirdNET confidence - Eurasian Jackdaw *(68 predictions)*	0.008	0.951
BirdNET confidence - Parakeet *(55 predictions)*	−0.131	0.286
BirdNET confidence - European Goldfinch *(40 predictions)*	−0.012	0.940
BirdNET confidence - Coal Tit *(15 predictions)*	0.103	0.715

and 12.3% (horizontal).

BirdNET confidence values and HARKBird localisations were not normally distributed. Friedman tests showed significant differences between the recording conditions for confidence values ($\chi^2(3)$=61.3, $p < 0.001$) and localisations ($\chi^2(3)$=24.0, $p < 0.001$). Dunn post-hoc tests revealed significant differences in confidence between the field recording and each VSE recording, and in localisations between the field and VSE vertical and horizontal recordings, and the VSE vertical and 45° recordings (Table 3.3).

There was a significant Spearman's rank correlation between device orientation and HARKBird localisations, but not between orientation and BirdNET confidence (Table 3.2).

BirdNET's precision and recall generally fell within similar ranges across recording conditions though there was some variation between conditions and confidence thresholds (Fig. 3.6). The metrics were limited or not calculable for Sites 3 and 4 due to a lack of avian activity. Precision was frequently over 0.8, sometimes limited by false positive predictions. Recall was always relatively low (below 0.3), but more consistent between recording conditions and confidence thresholds. Field recordings often had high precision but the lowest recall due to fewer BirdNET predictions.

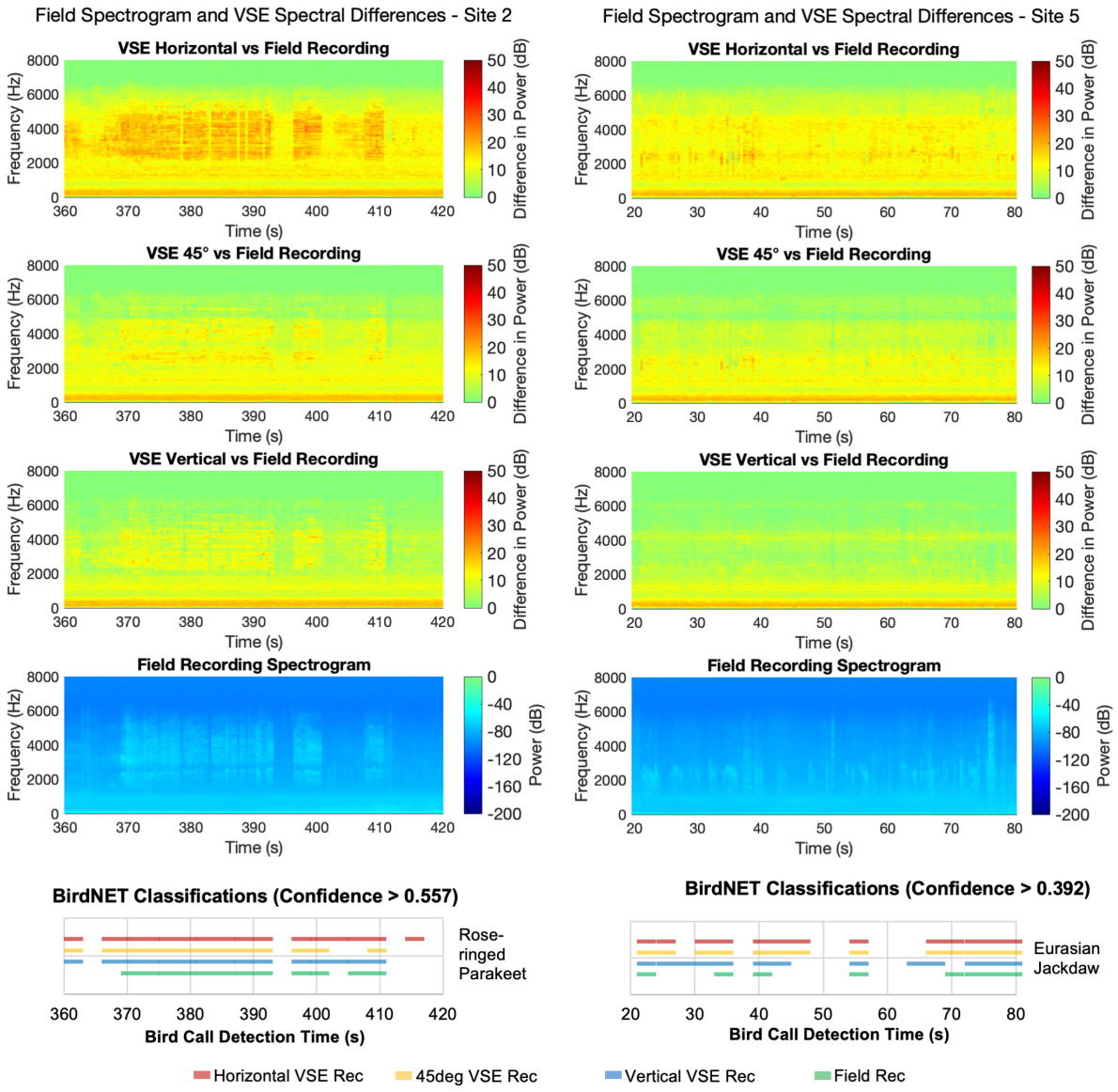


Figure 3.5: Field recording spectrogram, VSE recordings' spectral differences to the field recording, and BirdNET predictions over time for a 60 s window in the recordings for Site 2 and Site 5. Both recording segments contain many avian calls, reflected in the increases in spectral power in the spectrograms. The BirdNET predictions mostly correctly align with the spectral patterns but there is some inconsistency in predictions between recording conditions.

Table 3.3: Results from Dunn post-hoc tests following Friedman tests assessing for significant differences in BirdNET (BN) confidence and HARKBird (HB) localisations between all recording conditions (field and Virtual Sound Environment (VSE)).

		Recording Condition			
		Field - Vertical	VSE - Vertical	VSE - 45°	VSE - Horizontal
Recording Condition	Field - Vertical	-	**BN: $<$ 0.001** **HB: $<$ 0.001**	**BN $<$ 0.001** HB: 0.670	**BN: $<$ 0.001** **HB: 0.035**
	VSE - Vertical	-	-	BN: 0.971 **HB: 0.006**	BN: 0.300 HB: 0.283
	VSE - 45°	-	-	-	BN: 0.829 HB: 0.670
	VSE - Horizontal	-	-	-	-
Median of BN conf. values		0.799	0.928	0.855	0.884
St. Dev. of BN conf. values		0.247	0.207	0.204	0.282
BN No. species predicted		4			
BN No. predictions		75			
Median of HB localisations		*185*	*331*	*260*	*265*
St. Dev. of HB localisations		*40.7*	*2.23*	*4.41*	*23.6*
HB No. localisations		*9*			

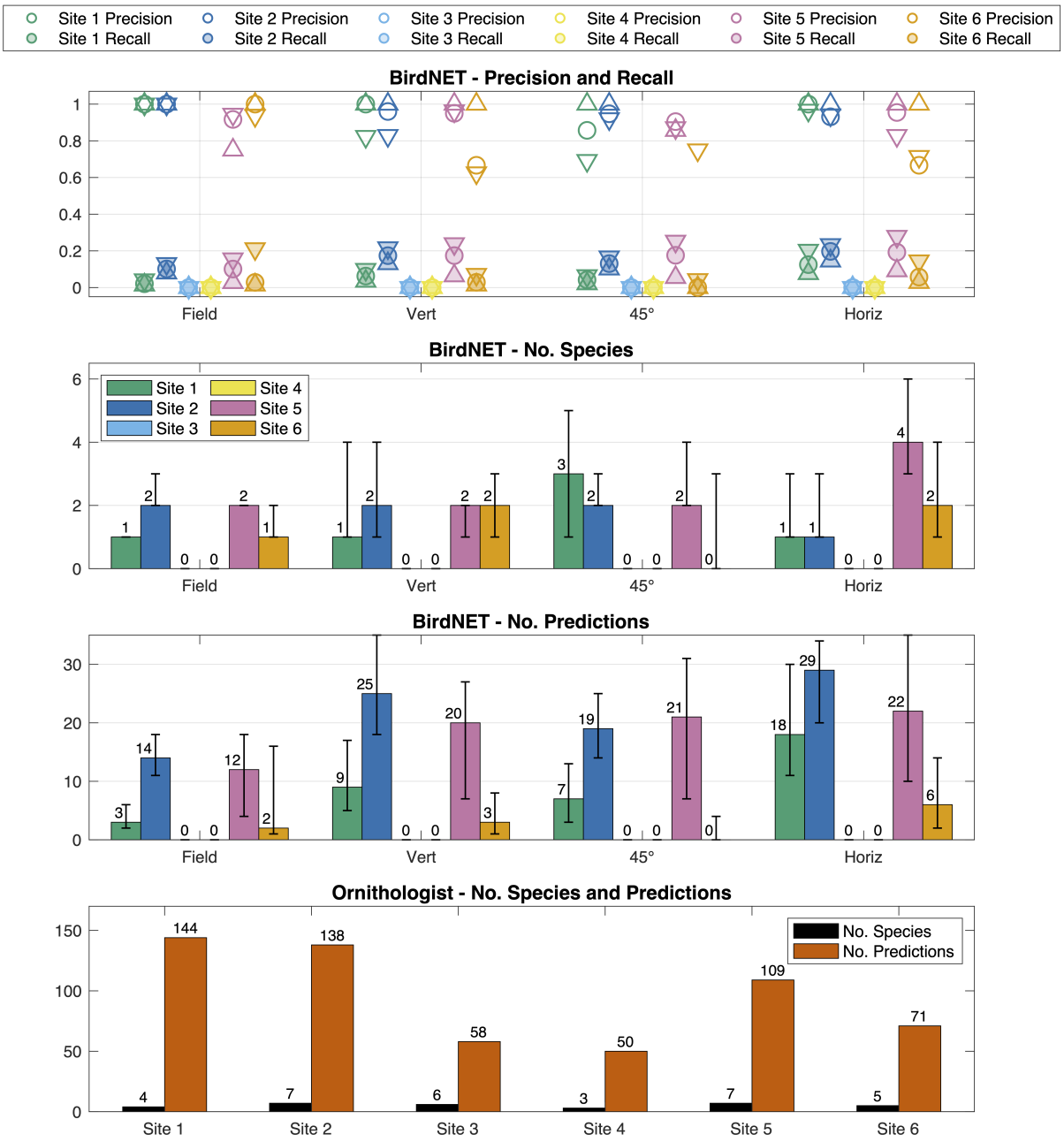


Figure 3.6: Precision and recall of BirdNET predictions on field recordings and VSE recordings at various orientations, filtered using minimum confidence thresholds of 0.25, 0.5 and 0.75. Precision and recall were calculated based on manual labelling of the omnidirectional channel of the ambisonic field recordings by a professional ornithologist. Values for precision when BirdNET made no predictions and recall when no bird calls were actually present are omitted to avoid division by zero. Downward facing triangles denote a confidence threshold of 0.25, circles 0.5 and upward facing triangles 0.75. Numbers denote the number of species and predictions (total number of 3 s windows with calls from each species) detected by the ornithologist and BirdNET (0.5 confidence threshold). Error bars indicate the number of species and predictions for the 0.25 (lower) and 0.75 (upper) confidence thresholds. Raw data and those for 0.01 and 0.9 confidence thresholds are presented in Table B.2.

3.3.2 BirdNET and HARKBird Performance on Additional Avian Calls

BirdNET correctly identified almost all added bird calls, with most over 0.7 confidence (Table 3.4). HARKBird localised only 30% of calls added by ambisonic encoding to within 45°, compared to 70% of those added from individual loudspeakers. Performance was worse for some species and the Blue Tit and Robin were sometimes not detected by BirdNET or HARKBird, respectively (Table 3.4). Calls added at higher elevations seem correlated with larger BirdNET confidence and HARKBird error (Fig. 3.7), though this is likely due to other factors for BirdNET.

We compared the BirdNET confidence values and HARKBird errors from both embedding methods using the Mann-Whitney U Test, as the data were not normally distributed. While the difference in BirdNET confidence was not significant ($p = 0.453$, $U = 26787$, $Z = -0.751$), HARKBird errors were significantly lower for the direct-from-loudspeaker bird calls ($p < 0.001$, $U = 25233$, $Z = -17.839$).

Table 3.4: Summary of BirdNET predictions and HARKBird localisations for bird calls added to various spatial soundscape recordings from Silwood Park reproduced through our ambisonic VSE. The added calls were embedded by either encoding to the ambisonic domain or playback from individual loudspeakers in the VSE. Abbreviations: Abs - Absolute; Az - Azimuth; BN - BirdNET; Conf - Confidence; deg - degrees; Err - Error; El - Elevation; HB - HARKBird. Positive azimuth errors are clockwise from the true value, and negative error values anticlockwise.

		Ambisonic Encoding			**Individual Loudspeaker**		
Site	**Common Name**	**Az, El (deg)**	**BN Conf**	**HB Az Err (deg)**	**Az, El (deg)**	**BN Conf**	**HB Az Err (deg)**
1	Robin	102, 56	0.97	166	126, 56	0.93	−51
1	Sparrow	182, 67	0.56	−160	198, 56	0.44	−118
2	Blue Tit	30, 45	0.4	10	18, 32	*N/A*	−16
2	Wood Pigeon	180, 45	0.73	−145	198, 56	0.9	10
3	Greenfinch	63, 35	0.93	−151	90, 56	0.99	1
3	Blackbird	260, 71	0.57	−145	270, 56	0.62	23
4	Blackbird	0, 0	0.71	35	0, 0	0.53	0
4	Blue Tit	150, 25	0.4	−35	162, 32	*N/A*	−22
5	Greenfinch	13, 68	0.9	92	342, 56	0.99	−18
5	Robin	243, 83	0.81	−133	270, 56	0.95	*N/A*
Mean (abs. values)			0.70	107	-	0.80	29
Standard Deviation (abs. values)			0.21	59		0.22	37

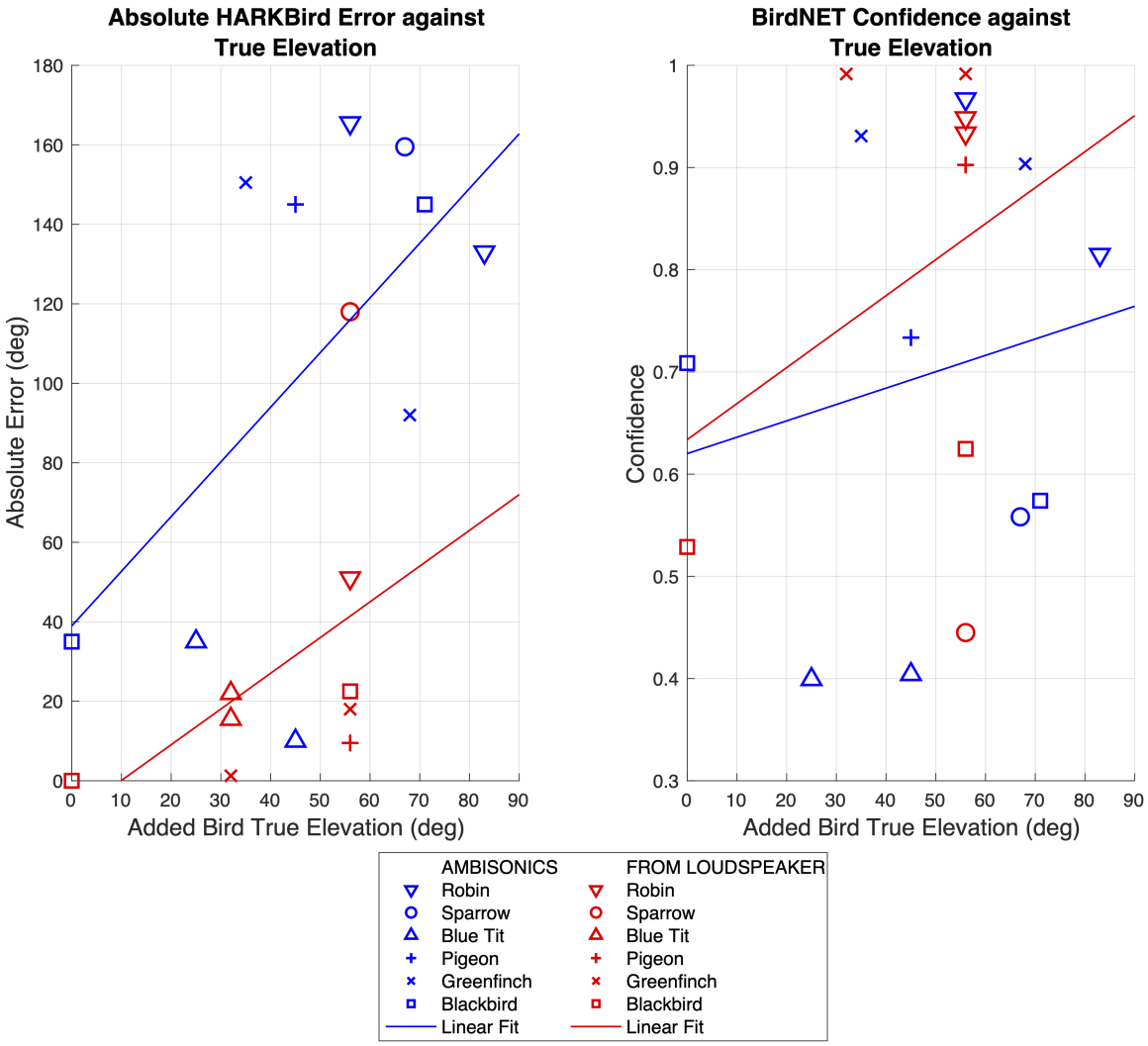


Figure 3.7: Graphs of HARKBird error and BirdNET confidence against the true elevation values of bird calls added to VSE-based spatial natural soundscapes via ambisonic encoding or playback from individual loudspeakers. Calls added from individual loudspeakers generally had both higher BirdNET confidence and lower HARKBird errors.

3.4 Discussion

Results from the spectral differences, VGGish, and, to a lesser degree, BirdNET, show potential for VSEs to simulate soundscapes well enough to test key aspects of passive acoustic monitoring technology. However, our results also highlight important caveats to this approach, including the limitations of ambisonics and the space housing our VSE.

3.4.1 Soundscape Reproduction Efficacy and Impact of Device Orientation

Spectral Differences

Though not a spatial consideration, it was fundamental to verify that overall spectral power patterns were preserved in the VSE (Fig. 3.3), as confirmed by the relatively low variance in spectral differences across all re-recordings (Table 3.1). This also suggests that level increases – though unexpected – were consistent within all re-recordings (and not, for instance, specific to transient sounds like bird calls). As such, level increases likely stemmed from factors that impacted the recordings overall, such as reverberation and standing waves from the untreated room initially housing the system.

In addition, the increases in spectral power as device orientation changed are most likely due to the larger number of loudspeakers in the horizontal plane of our ambisonic VSE. This stems from ambisonics' intended purpose of reproducing spatial soundscapes in a perceptually accurate manner, as humans are particularly sensitive to localising sounds horizontally.

VGGish Features and Acoustic Indices

We reiterate that while it is challenging to ecologically interpret the extracted features/indices [159], differences for an ideal VSE should be near zero. Acoustic indices' large and variable differences could stem from the limited recording duration and/or window length. These patterns may also reflect acoustic indices' inherent limitations: compared to VGGish features, acoustic indices have been found to be less robust to audio compression [176] and worse ecological descriptors when used as the input to various ecosystem classification tasks [123]. Conversely, VGGish features could also be more robust to ambisonic reproduction as they originate from a CNN trained on human-labelled audio [163,267] and may thus be biased towards human perception.

BirdNET and HARKBird

As BirdNET uses monophonic recordings, we anticipated most field recordings' species predictions to be detected in the VSE recordings. However, we expected higher and more consistent overlap. The 40.4-55.4% overlap range may relate to BirdNET's relatively low

recall on our recordings, with the large number of missed calls sometimes resulting in species predictions at different time-stamps in the field and VSE recordings. BirdNET also tends to return more predictions and higher confidence values on louder bird calls. This may explain the higher percentage overlap for the horizontal recordings, which were loudest due to the concentration of loudspeakers in the VSE's horizontal plane (Fig. 3.3). This increase in level also likely underpins the significantly higher confidence values for the VSE recordings, while BirdNET's lack of spatial considerations is reflected in the absence of both significant differences in confidence between re-recordings themselves (Table 3.3) and significant correlations between confidence and device orientation (Table 3.2).

For HARKBird, we expected the significant correlation between localisations and device orientation (Table 3.2), but it was surprising that so few HARKBird localisations in the VSE recordings matched the field recordings, and that the largest overlaps were observed for the 45° and horizontal VSE recordings. Given this and the post-hoc results (Table 3.3), it seems sound-source direction can be inaccurate in our ambisonic playback system and/or MAARU can suffer from considerable spatial aliasing between microphones. Indeed, on average 28.4% of HARKBird field localisations appeared in the vertical re-recordings, but most had >45° error. Moreover, the underlying MUSIC algorithm HARKBird uses to localise calls typically performs poorly on correlated signals [55]. Ambisonic playback's spreading of sound-source position through the activation of multiple loudspeakers [57] results in correlated, coherent signals originating from multiple locations, which likely further degraded HARKBird's performance.

Overall, we met aspects of Sub-projects 1 (Efficacy of Soundscape Reproduction) and 2 (Impact of Device Orientation), given the VSE's preservation of spectral patterns (despite overall level increases), VGGish features' minimal differences between the field and VSE, and the ability to use the VSE to draw baseline conclusions on device orientation's impact on BirdNET and HARKBird. In particular, we found that changing MAARU's orientation can progressively alter bird call direction of arrival estimation in HARKBird and can increase the level of sounds which are louder in certain directions (like those in the horizontal plane of our VSE). Such level increases can cause differences in BirdNET's performance as device orientation changes (mainly in the number of predictions and precision), though the confidence of overlapping predictions made at different orientations does not appear to significantly alter. However, we also identified limitations to both

HARKBird and our VSE. Future work should thus explore other, more accurate spatial audio reproduction techniques, such as those based on sound field reconstruction rather than human perception.

3.4.2 BirdNET and HARKBird Performance on Additional Avian Calls

Our tests towards Sub-project 3 revealed strong performance from BirdNET and potential for HARKBird on clear, highly-directional signals.

We note that all added bird calls were relatively loud compared to others in the soundscapes. Since a large range of SPLs could occur in the field depending on a bird's proximity to the microphone, we set added calls' gains to a level deemed perceptually appropriate for a nearby bird. This, however, is likely the main reason for BirdNET's high confidence values. Further work could explore adding calls at progressively lower levels. Nonetheless, these tests reiterated BirdNET's robustness to spatial variation given predictions' confidence values were not significantly different when added via ambisonic encoding or single loudspeaker playback, though it is unclear why the added Blue Tit vocalisations were not detected when reproduced directly from the loudspeaker.

The superior accuracy of HARKBird localisations on bird calls added from single loudspeakers further highlights the issue of spatial aliasing in ambisonic playback, with multiple loudspeakers activating for directional calls in our third-order VSE, and the difficulties that the MUSIC algorithm underlying HARKBird faces under such conditions [53, 55]. However, even with single loudspeaker playback, most errors were larger than in other testing of MAARU, in which calls have been localised to within ±10° [49]. This is most likely due to the addition of reverberation in both embedding methods blurring the source position of the bird calls, as it was generated from the impulse response at the closest of four azimuths (0°, 90°, 180° and 270°) to each added call's direct path, and could therefore differ by up to ±45°. Moreover, as the reverberation was generated in the ambisonic domain, it will have been reproduced from multiple loudspeakers, resulting in lower directionality.

Further, in previous experiments with MAARU, sounds were all reproduced in the horizontal plane. However, as reflected in Fig. 3.7, some error is expected as elevation increases, since sound sources are no longer in the same plane as MAARU's microphones

and may thus appear to originate from different positions.

While further work could interrogate the above findings, the ability to generate these insights indicates promise for using a VSE to evaluate the performance of passive acoustic monitoring software. Further, in totality our results suggest potential for using VSEs in other ecological applications requiring soundscape reproduction. For example, VSEs could provide a more realistic and immersive sonic environment for lab-based playback experiments to measure animal behaviour [58].

3.5 Guidelines and Recommendations

In this proof-of-concept, we have showcased the potential for VSEs to be used as test platforms for spatial and single-channel passive acoustic monitoring hardware and software. Further, we have highlighted the benefits and pitfalls of ambisonic sound field reconstruction in such VSEs. Based on this, we propose the following Guidelines and Recommendations for the use and further development of such platforms:

- **Guideline 1.** With careful experimental design, VSEs can be used to test monophonic passive acoustic monitoring hardware and software tools, and show potential for testing spatial passive acoustic monitoring technologies.
  - **Recommendation 1.** We hope a standard suite of analyses may be adopted for testing different classes of passive acoustic monitoring technologies, e.g., spectral differences and, as in [176], a set of typical acoustic indices alongside embedding into an audio-specific machine learning model.
- **Guideline 2.** Ambisonics is a useful and increasingly accessible method for reproducing real, spatial natural soundscapes in a VSE, thanks to its ability to capture actual sound fields and its setup-agnostic nature. Care should be taken, however, if using an ambisonic VSE to test spatial passive acoustic monitoring technologies, due to the potential for inaccurate sound field reconstruction arising from ambisonics' use of multiple loudspeakers and its bias to human perception.
  - **Recommendation 2.** We recommend considering ambisonics' limitations, using higher-order ambisonics (third order and above), and/or conducting analyses focussed on lower frequencies so as to have a larger sweet spot of

more accurate sound field reconstruction. Alternatively, we suggest exploring other methods of sound field reconstruction (e.g., vector-based amplitude panning [268]).

- **Guideline 3.** VSEs should ideally be housed in an acoustically-treated, anechoic space. However, ambisonic VSEs can be designed and built in a scalable, resource-efficient way.
  - **Recommendation 3.** We recommend using affordable or existing spaces and equipment for the creation of further, diverse VSEs, providing this permits for sufficiently accurate sound field reconstruction (e.g., sufficient number and spacing of loudspeakers for ambisonic systems) and the VSE is carefully calibrated. Doing so, VSEs can be constructed in days (weeks if sourcing most parts), while hours of field recording can allow for hours to days of lab re-recordings, depending on experimental conditions. Therefore, ecoacoustic VSEs can be built and piloted in a month. We hope this facilitates the development of more VSE test platforms for passive acoustic monitoring technologies across manifold spaces and budgets, while reducing the potential environmental impact of this work.

Chapter 4

Spatially segmenting soundscapes with a refined multichannel autonomous recorder for enhanced biodiversity assessment

Summary

1. Beamforming processes audio from multichannel recorders with small, integrated microphone arrays to spatially filter soundscapes, by amplifying sounds from target directions and attenuating ambient noise. It can thus improve signal-to-noise ratio, separate sources from overlapping sounds from other directions, and enable multichannel recorders to act like directional microphones that can be steered to different orientations after recording. Beamformed audio also more closely resembles the focal recordings of public sound repositories on which many machine learning bioacoustics classifiers have been trained.

2. However, due to practical barriers such as high power demands, large data needs, and ingress protection challenges, there have been few long-term deployments of autonomous multichannel recorders (i.e., that can record and transmit data without manual intervention). Further, beamforming has yet to be integrated into fully automated passive acoustic monitoring

pipelines.

3. To realise enhanced autonomous biodiversity assessment involving acoustic beamforming, we redesign the Multichannel Acoustic Autonomous Recording Unit's weatherproofing for improved robustness to long-term deployment, including damage from insects and animals, and assess its theoretical and real-world capacity to directionally-distinguish sounds with a lightweight, scalable beamforming technique. We iteratively refine and deploy MAARU in temperate and tropical forests in the UK, Brazil, USA and Malaysia and explore how various strategies for time-delay beamforming to fixed directions impact the outputs of BirdNET.

4. We find that beamforming results in significantly higher BirdNET prediction confidence scores than for concurrent predictions in omnidirectional audio, and it typically yields more predictions overall. We also show potential for using beamforming to detect habitat structure and distinguish bird communities in neighbouring areas of distinct land use. While further work is needed to integrate beamforming into large-scale, automated passive acoustic monitoring, its potential for spatially segmenting soundscapes promises to improve the accuracy and efficiency of autonomous ecoacoustic monitoring in diverse ecosystems.

4.1 Introduction

In earlier chapters, we advocated for the power of passive acoustic monitoring to surveil ecosystem health, biodiversity and resilience. We also introduced the concept of multichannel monitoring to capture spatial audio of natural environments and highlighted some of the benefits it offers for surveying ecosystems. In this chapter, we turn to a major challenge in autonomous passive acoustic monitoring and ways that multichannel recording can tackle it: disentangling the complex and often muddied mixture of sounds that make up soundscapes. Where acoustics has the unique advantage of hearing the unseen, it faces its own form of obstruction from overlapping sounds 'masking' target sound sources. Sufficiently loud noise, calls from other species and even multiple overlapping calls from a target species can all pose bottlenecks to acoustically surveying biodiversity. However,

such conditions are common in passive acoustic recordings, given they still often capture soundscapes with a single-channel omnidirectional microphone. Additionally, such recordings present a well-documented domain shift for machine learning bioacoustic classifiers, as they are usually trained on high signal-to-noise ratio focal recordings of animals, often captured with directional microphones and/or in close range of the emitter [169–171].

Some prior work has explored separating sound sources within single-channel recordings, for instance by decomposing spectrograms into components or basis functions (e.g., via principal component analysis) [269], or by using deep learning techniques [270, 271]. By identifying sounds' direction of arrival [42], multichannel monitoring offers a promising alternative approach, enabling target sound-producing individuals to be distinguished from overlapping sounds ('maskers') of different source location. Further, multichannel audio can be processed to reduce diffuse noise[1] and amplify sounds from target directions via a class of spatial filtering techniques known as 'beamforming'. These techniques generally involve taking the weighted sum of each channel's recording so that correlated (i.e., predominantly in-phase) signals like bird calls from a unique direction constructively superpose and decorrelated diffuse noise is attenuated by destructive interference. Beamforming therefore enables multichannel recorders to act as directional microphones that can be steered towards different 'look directions' after the fact of recording. Moreover, beamforming can be used to discern or even separate target signals from the acoustic mixture of soundscapes – though as we will discuss in later sections, beamformed audio is not necessarily as directional as that from a specially-constructed directional microphone. Nonetheless, beamformed audio stands to improve deep learning classifiers' accuracy by more closely resembling the focal recordings they were trained on.

Beamforming has been extensively studied for the design of sensor and antenna arrays with a wide variety of applications [272]. Its most simple implementation for planar waves[2] is 'delay-and-sum beamforming'. This method applies relative delays to each channel based on the difference in arrival time at each sensor of a signal from a given look direction, thereby aligning correlated signals from that direction across channels prior to summing them. Delay-and-sum beamforming is classed as a 'static' beamforming method, as it uses fixed weights for summing sensor outputs to beamform a given look direction.

[1] Diffuse sounds or noise are those which have energy spread evenly in all directions, such as the stridulation sounds of cicadas that are dispersed reasonably homogeneously across a forest.

[2] Planar waves suitably approximate acoustic waves under the far-field assumption.

In contrast, numerous 'adaptive' beamforming techniques have been developed, which dynamically compute optimal weights based on various features of a recording, such as the signal-to-noise ratio [272]. For example, minimum variance distortionless response ('MVDR') beamforming seeks to minimise variance (associated with noise) while preserving the target signal in the look direction with unity gain, avoiding the distortion that can occur when target signals are overly amplified by constructive addition [171, 273]. Such techniques are, however, more computationally intensive than static approaches. Importantly, the term 'adaptive' can also refer to beamforming methods that adapt their look direction towards target sources based on initial direction of arrival estimation. To avoid confusion, we shall therefore refer to beamforming methods that dynamically optimise their weights for a given direction as 'optimum beamforming' (as in [272]), and to those that adapt their look direction as 'directionally adaptive' and those that do not as 'directionally static'.

In acoustics, beamforming has particularly been applied to isolating speech, namely for hearing aid design [47] and teleconferencing [274], as well as robot audition [54]. Beamforming has also been explored in a number of bioacoustics studies. For example, Jones and Ratnam [273] used an array of four omnidirectional microphones at the vertices of a 3 m long square to recover individual anuran calls in choruses. Cai et al. [275] beamformed towards bird calls using the custom 'VoxNet' multichannel recorder [48], while Kojima et al. [276] separated bird calls using a small, commercially-produced seven-microphone recorder and a beamforming-derived spatial filtering technique in the robot audition software *HARK*. Wijers et al. [229] used time-delay beamforming to enhance target signals as part of a broader pipeline for identifying and localising various mammals and gunshots (indicative of poaching) via a number of custom four-microphone 'CARACAL' recorders in Zimbabwe. Time delays were estimated by cross-correlating signals from pairs of CARACAL's microphones and were also used to estimate sources' direction of arrival to refine the multi-recorder localisation process. In marine bioacoustics, recent work has used readings of acoustic pressure at numerous points along undersea fibre-optic telecommunication cables — effectively emulating a large hydrophone array — to localise and beamform towards whales [51].

However, beamforming has yet to be integrated into pipelines for long-term, fully autonomous, real-time acoustic biodiversity monitoring. The primary barrier to doing so

is that most multichannel recorders are not designed for fully autonomous deployments, with self-sufficient powering and data transmission coupled with robustness to prolonged environmental exposure. Consider for example the CARACAL recorder [229], which is one of the few multichannel recorders designed for passive acoustic monitoring research. CARACAL is small and low-cost (£150 in 2019), with an array of four MEMS microphones. However, its existing enclosure – a perforated plastic case sheltered by a large acrylic disk overhead – is not fully weather-sealed and the recorder relies on manual data retrieval from an onboard SD card.

In addition, most applications of beamforming in bioacoustics have to-date used a directionally adaptive approach, employing a number of different direction of arrival estimation techniques to identify look directions for target sources. This offers a more complete picture of the distribution of callers around an array, enabling, for instance, estimates of species abundance and movement. However, directionally adaptive beamforming can be relatively processing-heavy, and many of the bioacoustic studies that employ it are taxonomically narrow – perhaps because monitoring more taxa would increase the number of look directions and overall processing load further. Also, as shown in the previous chapter, azimuth direction of arrival estimates with planar arrays – which are more typical of small, low-power arrays suited to passive acoustic monitoring [49, 229, 277] – are sensitive to deployment orientation and often erroneous for sources at large elevation angles relative to the array.

Directionally static beamforming may be able to address similar ecological questions – and other relevant ones – and improve signal-to-noise ratio with a less processing-heavy approach that could be more readily incorporated into fully automated monitoring workflows. While such gains in processing efficiency may be negligible for some autonomous workflows that stream audio to a server for analysis, they may still be notable for particularly large-scale or taxonomically-broad surveys and will be especially important for eventually running processing on-device, as prior work has called for [229]. Further, beamforming on-device to a handful of carefully-chosen, validated preset directions may reduce the amount of data generated, lowering data storage costs and the power required for wireless data transmission, which might compensate for the increased energy demand of local processing.

Finally, directionally static beamforming may be well-suited to address certain eco-

acoustic questions that more closely interrogate the spatial features of natural soundscapes in relation to the landscapes that give rise to them [31]. By using directionally static beamforming to focus on sounds from different prevailing directions, it may be possible to 'segment' soundscapes in a manner that uncovers habitat structure, distinguishes distinct neighbouring land use and reveals the local spatial heterogeneity of biodiversity over areas of tens to thousands[3] of square meters. With ideal hardware and recording conditions, segmenting soundscapes via beamforming might even yield insights into beta-diversity – an increasingly prevalent topic of ecological research since the early 2000s [278].

In this chapter, we contribute to tackling these barriers to beamforming in fully-automated acoustic biodiversity monitoring pipelines by iteratively refining, deploying and beamforming with MAARU [49] in the UK, USA, Brazil and Malaysia. We first detail substantial revisions to MAARU's weatherproofing designed to improve its robustness to long-term environmental exposure, which we test in the UK, Brazil and Malaysia (**Sub-Project 1** – Section 4.2). Next, we evaluate how MAARU's array configuration shapes its beamforming capacity by analysing its theoretical power response pattern with and without a steering vector applied (**Sub-Project 2** – Section 4.3). We examine these patterns with subsets of MAARU's microphones to consider scenarios where certain microphones fail. We then detail a preliminary investigation into how directionally static beamforming can improve the confidence score of avian deep learning classifier BirdNET [167] in the UK and Brazil (**Sub-Project 3** – 4.4). Finally, we explore MAARU recordings from the UK, USA and Malaysia to test various strategies for segmenting soundscapes, such as beamforming away from sources of noise and/or towards directions of interest (**Sub-Project 4** – Section 4.5). For some sites, we beamform towards neighbouring areas of distinct land use to determine whether multichannel recorders like MAARU could monitor ecological gradients with fewer recorders than sampling sites. We again evaluate these beamforming strategies in terms of BirdNET predictions, given its widespread use, existing application to global biodiversity monitoring [170, 172] and because it was trained on 100,000s predominantly focal recordings (including of some non-avian classes like anurans, which we study at one of our US sites).

We find that despite the limitations to our beamforming method and hardware, our approach to spatially segmenting soundscapes can significantly improve the confidence

[3]For particularly loud sounds and/or in more open environments.

score of BirdNET predictions (associated with higher prediction accuracy [184]), returns more predictions, and can infer spatial heterogeneity in biodiversity potentially indicative of landscape and/or land use patterns. We therefore emphasise how our lightweight beamforming approach can address important ecological questions while providing a readily-achievable route to enhancing real-time automated acoustic biodiversity monitoring with spatial insights.

4.2 Sub-Project 1: MAARU weatherproofing redesign and field robustness tests

4.2.1 Methods

Weatherproofing Redesign

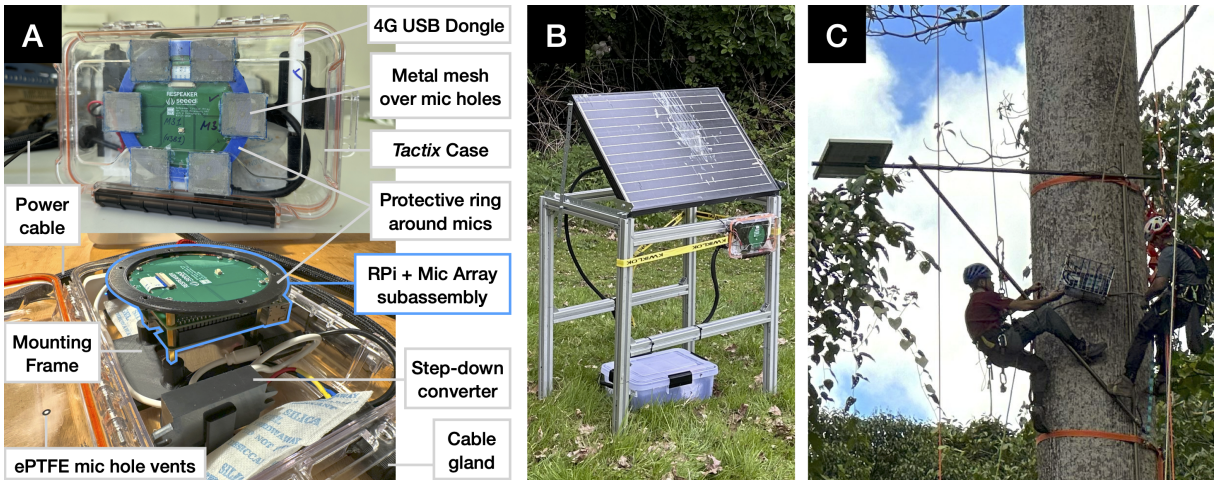


Figure 4.1: **A:** Redesigned MAARU weatherproofing enclosure. **B:** MAARU deployment in the UK. **C:** Professional tree climbers Jamiluddin Jami and Louis Makhluf mount a MAARU to a tree in Malaysian Borneo.

MAARU's original weatherproofing was based on a modified off-the-shelf plastic container [205]. This was intended to be highly affordable and require no specialist tools to build. Early deployments revealed robustness challenges, though, in particular from animal interference, such as rodents chewing on power cables, and impacts, such as falling from height. A custom 3D-printed enclosure for semi-autonomous deployment (intended for sites without mobile signal) was designed to address some of these issues with a sturdier enclosure. However, it was cumbersome to print (>15 hr on a *Prusa* MK3S) and required

applying a coat of resin to fully seal against water, limiting accessibility for users without access to a well-equipped workshop. The water-resistant ePTFE (expanded polytetrafluoroethylene) vents used to cover the microphone holes in both original designs were also found to be susceptible to insect ingress – specifically, termites chewing through the vents – during deployments in the Brazilian Amazon.

For the redesigned weatherproofing, we thus opted for a durable polycarbonate, water-resistant case (*Tactix* Water Proof Case (L), ∼10–15 USD depending on vendor) to house the core electronics (Raspberry Pi, sound card, microphone array, 4G dongle and a 12–5 V step-down converter). The case lid hinges for easy access to the internals. To prevent insect ingress, we used much smaller microphone holes and ePTFE vents, as the MEMS microphone ports are <0.5 mm in diameter. We also attached patches of fine stainless steel woven mesh (100 mesh, 0.15 mm hole size) over each microphone hole on the exterior of the case, using a water- and UV-resistant adhesive (Fig. 4.1A).

To accurately align the microphone ports with the corresponding holes in the case, we designed a 3D-printed mounting frame that sits near-flush against the base and sides of the case. The frame attaches to the case via velcro and the electronics mount to it without adhesives, allowing all parts to be repaired or replaced with ease. The Raspberry Pi, sound card and microphone array form a sub-assembly, joined by screws and metal standoffs, that rests on springs glued into the mounting frame. A 3D-printed ring also sits around the microphones and flush with the top of the case. It contains a small hole for each microphone port and acts as an additional physical barrier to insect ingress. The springs prevent excessive force being placed on this sub-assembly when the case lid is closed.

A single power cable enters the case via a watertight cable gland. This can either be a USB cable connected to the Raspberry Pi, or, for solar powered deployments, a two-core power cable (American Wire Gauge (AWG) 17 was suitable for our powering setup), which connects to the step-down converter via bullet-style crimp connectors. Wherever possible, appropriately specified crimp connectors were used instead of soldering to make electrical connections, as they can be easily dis/re-connected. The holes for the microphones and cable gland are all manually drilled and accurately located using 3D-printed jigs that can attach to the *Tactix* case. The electronics power cable's other end attaches to a 30 A solar charge controller which is housed with a lithium iron phosphate (LiFePO4) battery in a

separate IP56-rated polypropylene box (supplier: *Solent Plastics*, product code: WBAT-S, ∼10 USD). Both the electronics power cable and a heavier duty power cable that runs to the solar panel (in our case, AWG 13 was suitable) enter this box via cable glands in manually-drilled holes. Where possible, we wrapped cables in a semi-rigid plastic cable tidy and nylon cable mesh to protect them from rodents and other animals. If budgets permit, we recommend using armoured cable for long-term deployments.

UK and Borneo Deployments

MAARUs were deployed at four temperate sites at Imperial College's Silwood Park campus (Ascot, UK) for various periods of time (from 14 to 125 days) between February and July 2024. MAARUs and 60 W solar panels were mounted to custom-built metal frames fixed to the ground, with the battery box placed underneath (Fig. 4.1B). In September 2024, MAARUs were deployed at six tropical sites spanning a land-use gradient in and around the Danum Valley Conservation Area in Sabah, Malaysia. Two devices were deployed in each of: old-growth forest, logged forest, and smallholder oil palm plantations. Data were collected between approx. two weeks and five months, depending on devices' survival time (see Section 4.2.2). For access to sunlight and protection from animals on the ground (e.g., elephants), the old-growth and logged forest devices were mounted to trees by professional tree climbers. Solar panels (three 55 W and one 100 W) were attached to triangular structures assembled from aluminium poles (Fig. 4.1C). MAARUs and battery boxes were placed in a cage built from six steel grills (supplier: *DAISO*, ∼6 USD per cage), to protect them from animals in the trees (e.g., monkeys). The cages and pole structures were fixed to trees with heavy-duty ratchet straps. For the oil palm sites, the device cage and 100 W solar panels were mounted to slotted angle bars inserted in the ground.

Across all sites, MAARUs were deployed near the edges of different habitats or land-use types and/or within audible range of roads or other sources of anthropogenic noise from prevailing directions, so as to explore various applications of beamforming. All recordings were made with a 16 kHz sampling rate, compressed to FLAC (Free Lossless Audio Codec) on-device and streamed over 4G to *Box* Cloud Storage via FTPS. For continuous 24/7 recording, this creates ∼8 GB/day of data (∼250 GB/month), so all devices had a minimum 300 GB/month cellular data plan.

4.2.2 Results

MAARU's redesigned weatherproofing improves robustness to previous failure modes including impacts, rodents chewing on cables, and insect ingress, while enabling better access to internals and easier manufacturing than the previous 3D-printed enclosure. However, new failure modes have been discovered. For the UK deployments, the 3D-printed parts were made from PLA (polylactic acid), which is safe (minimal fumes) and straightforward to print, but has a fairly low glass transition temperature (T_g) of ∼57°C [279]. This was expected to be sufficient, however parts in three units warped due to heat from the electronics, damaging several microphones in these units (sometimes in just two weeks after deployment). For the Borneo deployments, these parts were printed from ABS (acrylonitrile butadiene styrene), which has a T_g of ∼107°C [279], and did not warp.

In Borneo, the old-growth and logged forest units failed after ∼2-4 weeks. This appears to be due to water ingress from heavy rainfall piercing the ePTFE microphone vents (not all units have been retrieved yet). All old-growth and logged forest units were deployed horizontally (i.e., with the ePTFE vents perpendicular to rainfall) to provide the freedom to beamform in any horizontal direction. MAARUs were deployed vertically in the UK and oil palm sites and did not face this issue. The oil palm devices were also sheltered by the solar panels.

Finally, for all deployments, the 55 W and 60 W panels did not provide enough power for continuous 24/7 monitoring (perhaps due to lower insolation in the UK and some obstruction from trees and clouds in Borneo), whereas the 100 W panels in Borneo did.

4.2.3 Discussion

With tweaks to MAARU's deployment set-up, its new weatherproofing design has potential to facilitate months-long deployments in both temperate and tropical climates.

While it was surprising that the 3D-printed PLA parts warped in the UK deployments, we note that the Raspberry Pi and 4G dongle can generate considerable heat, especially when the dongle has poor signal or there is a backlog of recordings to upload (after e.g., a network outage or lack of power overnight). The heat from these components was likely exacerbated by them being housed in a clear, sealed enclosure exposed to sunlight – though we note one of the UK devices which did not run during deployment (owing to a faulty step-down converter) was recovered intact after being in the field for the

same period as the devices that experienced part-warping. This points to the electronics being the primary problematic source of heat. Since switching the 3D-printed parts to ABS avoided this issue for the Malaysia deployments, we recommend using it or other materials with a similar or higher T_g in future.

To prevent water ingress, we recommend either sheltering any units deployed horizontally (e.g., with the solar panel or some other cover as in [229]) or deploying the MAARU units upside down, which should allow for similar beamforming and localisation capabilities. Lastly, if continuous, long-term monitoring is required, we recommend using 100 W solar panels, as limited sunlight can cause intermittent recording/uploading with less powerful panels.

Despite the device failures encountered, sufficient data was gathered to investigate beamforming in the subsequent sections. Moreover, for the UK deployments where part-warping caused individual microphones on several units to fail, there is also an opportunity to evaluate MAARU's beamforming capacity with subsets of fewer microphones. We examine this from a theoretical perspective in the following section.

4.3 Sub-Project 2: Theoretical Array Directivity

4.3.1 Background on Array Geometry and Directivity

A microphone array's capacity to preferentially receive acoustic power from different directions ('directivity') and accurately beamform towards them depends on trade-offs between array size, geometry, number of elements and target frequencies. Beamforming steers the directional power pattern of an array, generally producing a main lobe oriented towards the specified look direction (Fig. 4.2). To do so, beamforming typically utilises integrated arrays in which microphones are closely spaced, since a larger inter-microphone distance lowers the frequency threshold above which spatial aliasing occurs. As highlighted in the previous chapter, spatial aliasing can lead to confusions in direction of arrival estimation. For beamforming, spatial aliasing causes sounds from non-target directions to appear in the beamformed signal. Just as temporally subsampling a signal will introduce unwanted 'grating' frequencies in the sampled signal's spectrum, spatially subsampling an acoustic field results in 'grating lobes' appearing alongside the main lobe in the directivity pattern of a steered array (Fig. 4.2). This frequency threshold is dictated by the spatial sampling

theorem [274], which states that the inter-microphone distance d must conform to:

$$d < \frac{\lambda_{min}}{2} \tag{4.1}$$

where λ_{min} is the minimum wavelength of sound that can be detected without aliasing for a given inter-microphone distance. Larger distances thus cause larger wavelengths (lower frequencies) to be impacted by spatial aliasing. However, for a given frequency and number of microphones, increasing the spacing produces a narrower (i.e., more directional) main lobe. For a set array size, increasing the number of microphones reduces the inter-microphone distance and increases the spatial sampling resolution. This decreases the size of side lobes in the directivity pattern at different frequencies, reducing sounds from directions other than the main lobe. Additionally, the main lobe is usually narrower for higher frequencies as they are more directional. Hence, the directivity of beamforming is generally more precise for higher-pitched sounds.

Finally, the symmetry of directivity patterns and an array's capacity to discriminate between opposing directions is also dictated by array geometry. For a linear array, directivity will be symmetrical on either side of the axis along which the microphones are placed [274]. Thus, even in two dimensions, a linear array cannot discriminate between sounds to either side of it. Similarly, while a planar array like MAARU's will be directional in two dimensions within the microphone plane, its 3D directivity will be symmetrical about this plane. This geometric constraint limits planar arrays' ability to discriminate between sounds above and below the microphone plane (Fig. 4.3), such that beamforming to a positive angle of elevation results in a second symmetrical main lobe oriented to the corresponding negative angle of elevation.

Note on Azimuth and Elevation Convention

Throughout this chapter, we always define azimuth and elevation angles relative to a coordinate system with its origin at the centre of MAARU's planar microphone array. We always consider the azimuth (x-y) plane as that which contains MAARU's microphones. Azimuth ranges from $-180°$ to $180°$ and is positive anticlockwise from the x-axis; elevation ranges from $-90°$ to $90°$ and is positive anticlockwise from the azimuth plane to the positive z-axis. When MAARU is deployed vertically, we refer to azimuth angles aligned with its x- and y-axes as left ($0°$), up ($90°$), right ($180°$) and down ($-90°$), while for

horizontal deployments we denote these azimuth angles using cardinal directions, with east corresponding to 0°.

4.3.2 Methods

We calculated the directional power pattern (Watts/steradian) for MAARU's full six-microphone array, as well as sub-arrays of three and four microphones, across several frequencies up to 8 kHz (as MAARU records at a 16 kHz sampling frequency by default). The power pattern at a reference distance across angular directions of azimuth θ and elevation ϕ is defined [280, 281] as:

$$|Af(\theta, \phi)\mathbf{w}^H\mathbf{s}|^2 \tag{4.2}$$

where A is the nominal field amplitude, $f(\theta, \phi)$ denotes the element field pattern, and $\mathbf{w}^H\mathbf{s}$ is the array factor composed of element weights $\mathbf{w}^H$ and steering vector $\mathbf{s}$. Patterns were computed using the 'phased.ConformalArray pattern' function in the *MATLAB* (version 2024a) Phased Array Toolbox. We defined the microphone arrays as 'phased.ConformalArray' objects consisting of various arrangements of identical omnidirectional microphones. The microphones were defined with 'phased.CustomMicrophoneElement', to which we passed the frequency response of the MEMS microphone used on MAARU's array (*Seeed Studio* ReSpeaker 6-microphone array [252], microphone: MSM321A3729H9CP, *MEMSensing Microsystems*, Suzhou, China); the microphone frequency response is relatively flat but increases from 2 kHz. As MAARU's array is planar, we primarily consider its 2D directivity in the azimuth plane. However, to exemplify the symmetry of MAARU's directivity in three dimensions, we also plot its 3D directivity at a reference frequency of 4 kHz.

4.3.3 Results

MAARU's array power patterns in the azimuth plane are plotted with and without a steering vector applied in Fig. 4.2. Across all plots, the patterns at lower frequencies (up to 1 kHz) are more isotropic, while higher frequencies exhibit much more directional patterns. Nonetheless, without any steering applied (i.e., beamforming with uniform weights; middle column of Fig. 4.2) MAARU still receives power from directions all around

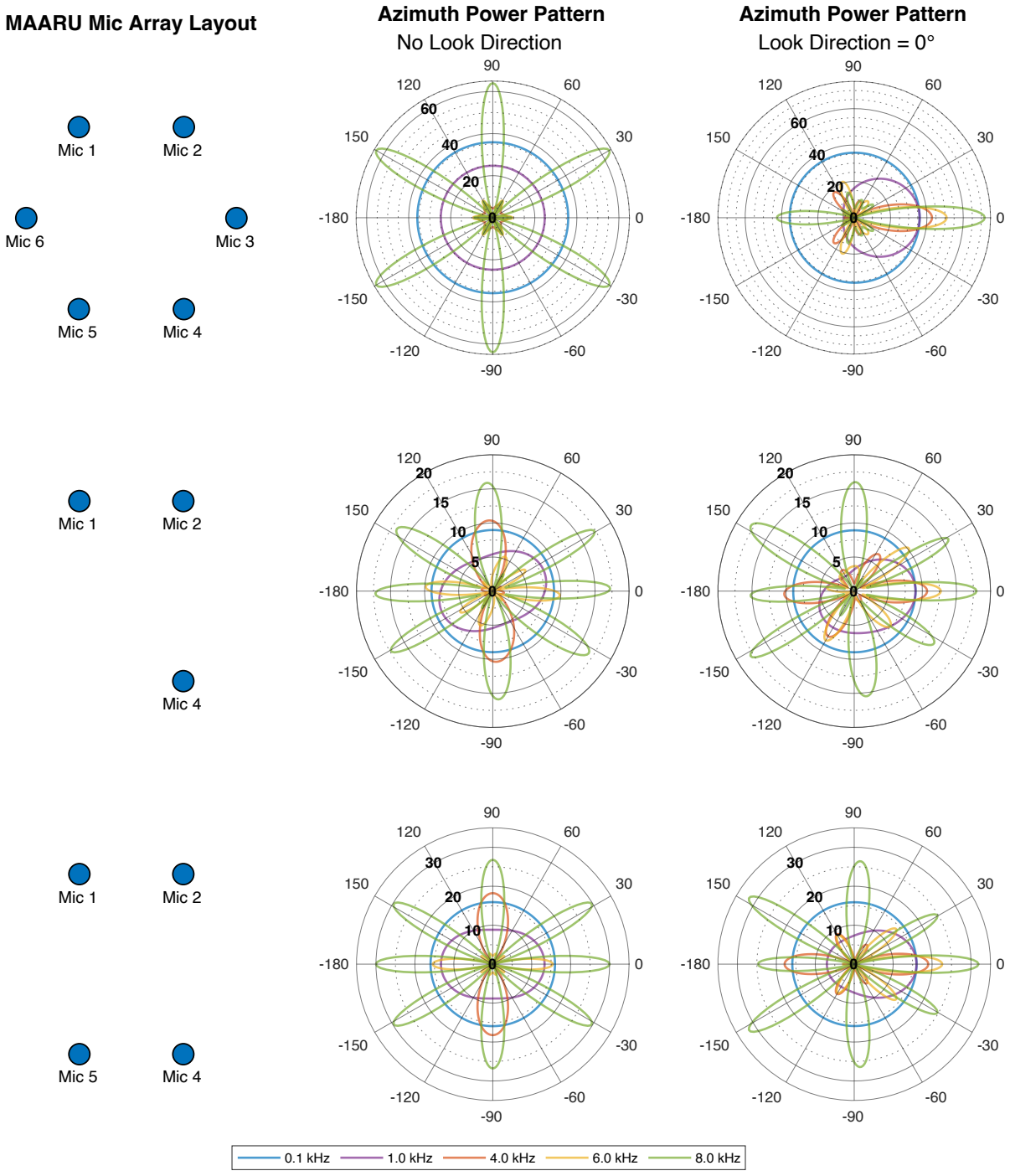


Figure 4.2: Plots of MAARU's microphone array power patterns at a reference distance (radial axis units: Watts/steradian) for all six microphones and subsets of three and four microphones shown in the left column. Middle column: beamforming weights equal to unity applied (akin to simple stacking of the channels without any beamforming look direction; Right column: weights corresponding to a delay-and-sum beamformer with a 0° azimuth look direction applied. Note the scale of the radial axis varies between subplots, as more power is concentrated in the 0° look direction when beamforming with all six microphones (upper right subplot).

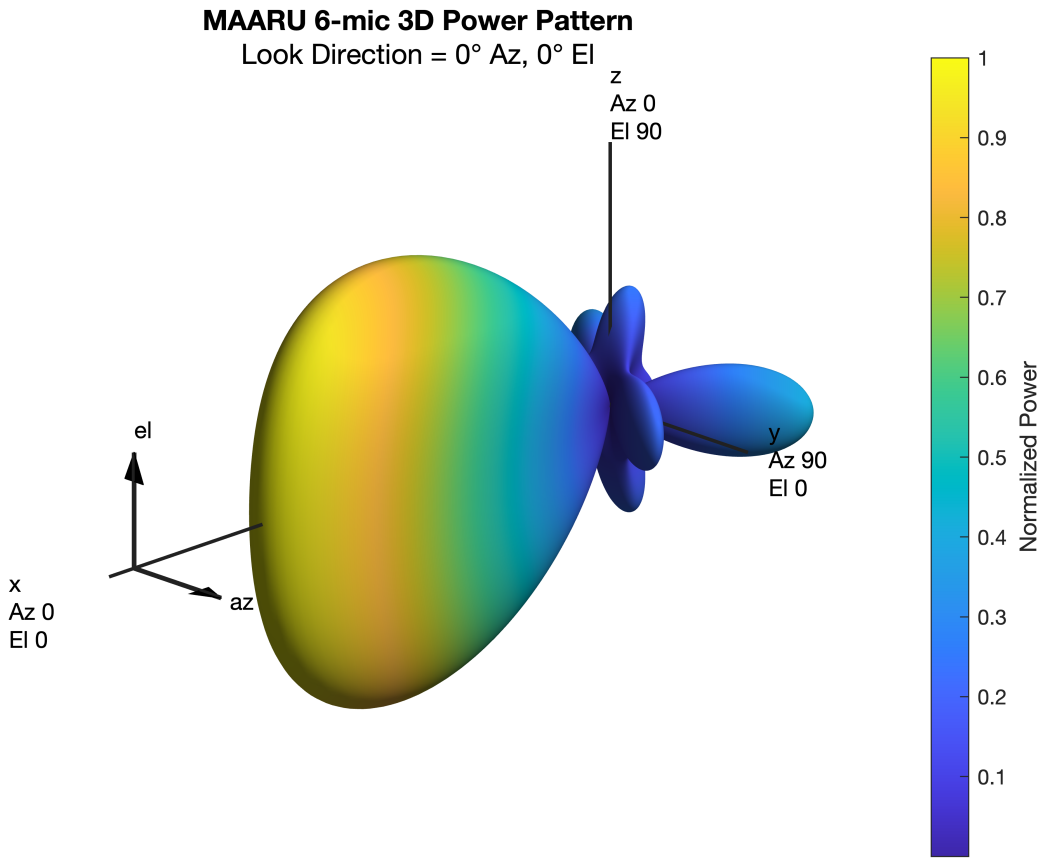


Figure 4.3: Three-dimensional plot of MAARU's normalised power pattern at a reference distance for 4 kHz signals, with a 0° azimuth look direction. Az: Azimuth, El: Elevation.

its array – albeit in the directions of a set number of lobes for higher frequencies. This approach should, however, still attenuate diffuse, uncorrelated noise. Steering the array to a look direction of 0° azimuth (rightmost column of Fig. 4.2) is reasonably effective when using the full six-microphone array, producing a clear main lobe that becomes increasingly narrow from frequencies of 1 kHz and above. However, as expected, side lobes are much larger and more numerous with fewer microphones. For the three-microphone configuration, the patterns with and without steering are also less symmetrical about the x (0° azimuth) and y (90° azimuth) axes.

The 3D power pattern is plotted in Fig. 4.3, confirming symmetry either side of the azimuth plane which contains the microphones.

4.3.4 Discussion

MAARU's directivity patterns reflect the expected theoretical considerations for array design, with greater directionality when beamforming with more microphones at higher fre-

quencies. While directivity only becomes more directional from 1 kHz and up, this should be suitable for targeting bird vocalisations, given many lie between 1 and 6 kHz [26,171]. Beamforming with fewer than all six of MAARU's microphones results in much lower directionality, suggesting it should be avoided except in cases where doing so is necessitated by e.g., failure of individual microphones. Despite this, beamforming with three or four microphones should still reduce diffuse noise and attenuate sounds from directions between lobes, while slightly amplifying sounds from the look direction. However, as directions with large side lobes will still be included – in particular for frequencies above 4 kHz impacted by spatial aliasing – beamforming with three or four microphones should primarily be used to improve signal-to-noise ratio rather than distinguish sounds from a particular direction.

4.4 Sub-Project 3: Initial Beamforming Validation

4.4.1 Methods

As a preliminary exploration into the potential benefit of beamforming with MAARU for BirdNET predictions, we applied beamforming to MAARU recordings from the UK and Brazilian Amazon in four arbitrary preset directions. We then compared the confidence scores of BirdNET predictions that were for the same species and shared the same timestamp in the omnidirectional recording from a single microphone on MAARU's array and beamformed audio from at least one of the preset directions ('matching predictions').

We applied this approach to 55.1 hours of acoustic data from Manicoré (Brazilian Amazon; May 2023) and 79 hours of data from Silwood Park (Ascot, UK; March 2023), both recorded at a 16 kHz sampling rate and 16-bit resolution. The recordings were captured semi-autonomously in April (UK) and May (Brazil) 2023 using four MAARUs enclosed in the fully 3D-printed variation of its weatherproofing and mounted vertically to trees. We employed delay-and-sum beamforming using the ODAS Python package (Open embedded Audition System [282]) to steer recordings in the following directions relative to MAARU's array: [Azimuth, Elevation] = [0°, 45°], [90°, 45°], [180°, 45°], [−90°, 45°]. We note that the +45° elevation angles used here will produce additional reflected main lobes at -45° elevation (i.e., behind the array; Section 4.3). However, as the devices were mounted to trees, we expect sounds from these negative elevation directions to have been

largely obstructed.

To test the impact of beamforming on BirdNET's confidence, we applied a Wilcoxon signed-rank test to confidence scores above 0.4 from matching predictions.[4] Since the relationship between confidence score and prediction accuracy varies by species [184], we also applied this test to the confidence scores of each individual species with over 30 matching predictions of confidence above 0.4 (14 species total). As this involved making multiple comparisons, a Bonferroni correction was applied to the species-specific tests, adjusting the alpha value to $0.05/14 = 0.00357$. In cases where a prediction appeared in multiple beamformed recordings, we retained that with the highest confidence score.

4.4.2 Results

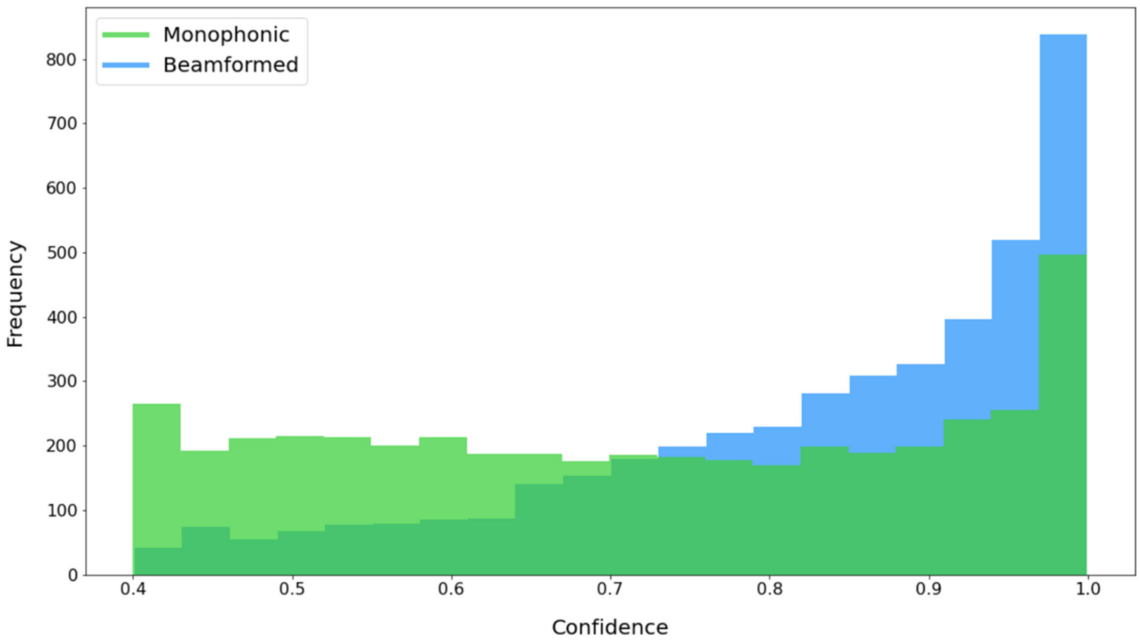


Figure 4.4: Histogram of BirdNET confidence scores of all matching predictions for both omnidirectional ('monophonic'; green bars) and beamformed (blue bars) recordings from the UK and Brazil as part of Sub-Project 3. Recordings beamformed to arbitrary preset directions result in higher BirdNET confidence scores than single-channel omnidirectional recordings.

Mean confidence using beamforming was higher than that of omnidirectional recordings (Fig. 4.4; beamforming: 0.83, SD= 0.15; omnidirectional: 0.71, SD=0.19). The median confidence for beamforming (0.872) was also higher than for omnidirectional recordings (0.718). The Wilcoxon signed-rank test confirmed this increase was significant for

[4]Slightly below a ballpark 0.5 confidence threshold for achieving a low false-positive rate, suggested based on findings from [283] in a 2023 review [170].

Table 4.1: Results of Wilcoxon signed-rank tests applied to BirdNET confidence scores of matching predictions from beamformed and omnidirectional audio for individual species in UK and Brazil recordings for Sub-Project 3. Species tested had at least 30 matching predictions of at least 0.4 confidence and all show a significant increase in confidence for beamformed recordings.

Species Name	**Sample Size**	**W-statistic**	**z-score**	***p*-value**
Mealy Parrot	107	5020	6.62	<0.001
Green-Backed Trogon	47	1092	6.52	<0.001
Tawny-bellied Screech Owl	752	240746	16.64	<0.001
Crested Owl	35	603	5.38	<0.001
Red-necked Woodpecker	30	439	4.75	<0.001
Eurasian Jackdaw	1263	724930	25.13	<0.001
Carrion Crow	146	10656	10.34	<0.001
Eurasian Nuthatch	447	94264	16.17	<0.001
Eurasian Green Woodpecker	214	20562	9.99	<0.001
Eurasian Blue Tit	171	12120	7.35	<0.001
Green Spotted Woodpecker	359	54227	11.14	<0.001
Great Tit	253	28703	10.85	<0.001
Stock Dove	69	2415	7.22	<0.001
Coal Tit	84	2564	3.47	<0.001

the amalgamated results across all species ($W = 8457060.0$, $z = 44.64$, $p < 0.001$) as well as for the species-specific results (Table 4.1).

While not tested here, simply stacking the 6 channels may also improve signal-to-noise ratio and yield increased BirdNET confidence with even more simple processing. However, as shown in Fig. 4.2, doing so will include sounds from a greater number of directions (in particular for the higher frequencies in which bird vocalisations typically lie), which is likely to limit the confidence gain compared to the more directional beamforming approach we adopted.

4.4.3 Discussion

This Sub-Project demonstrates potential for using beamforming with MAARU to improve avian biodiversity monitoring with BirdNET. We note that while confidence scores are not probabilities, a higher confidence has generally been found to correlate with higher prediction accuracy for a wide range (over one thousand) of species [184]. This improvement is likely down to the beamformed audio featuring a higher signal-to-noise ratio and more closely resembling focal recordings.

In addition, while retaining the highest confidence score for predictions that appeared

simultaneously in multiple streams of beamformed audio may appear to have biased these results in favour of beamforming, we note that this approach bears some similarity to methods that use multidirectional beamforming to identify source directions of arrival prior to then beamforming in such directions to separate the sources [273,275,276]. Here, the number of look directions is much reduced (only four), and the confidence score is a proxy for determining the direction that is likely closest to the source direction.

4.5 Sub-Project 4: Beamforming in Directions of Ecological Interest

In this final Sub-Project, we aimed to use knowledge of habitat and site structure to explore whether the benefits of beamforming established in the previous section could scale to more ecologically relevant applications. Specifically, we investigate BirdNET predictions from audio beamformed away from directional sources of noise ('undesired classes' like 'engine'), towards sources of biophony ('target' classes), and towards neighbouring but distinct sites along a land-use gradient.

4.5.1 Methods

We examined recordings from single days (up to 24 hr) within the deployments at the sites in the UK and Malaysia described in Section 4.2.1, and two short-term semi-autonomous recordings from Ithaca, NY, USA. The first Ithaca recording lasted ∼40 mins and was taken on 14 May 2024 outside the Cornell Lab of Ornithology (US Site 1), facing Sapsucker Woods pond, which a flock of Canada Geese were wandering in and around. In this deployment, a MAARU was mounted horizontally to a tripod without any weatherproofing. The second is a ∼3 hr recording from 14 June 2024 taken with MAARU in the weatherproofing described in Section 4.2.1 (minus the outer metal mesh over the microphone holes), attached vertically to a tree next to a path and small pond deeper within the Sapsucker Woods (US Site 2). Green Frogs (*Lithobates clamitans*) were present and vocally active in this small pond (and potentially around it). In both US recordings, a 16 kHz sampling rate and on-device compression to FLAC were used.

For all recordings, we performed time-delay beamforming using the 'Phased Array toolbox' in *MATLAB* (version 2024a). We used all six of MAARU's microphones for

all sites except UK Site 2, where two microphones were damaged prior to deployment. For this site, we explored beamforming with the subsets of three and four microphones analysed in Section 4.3. For each site, we steered the array in one or more directions to explore various applications of beamforming, as detailed in Table 4.2. We then passed the beamformed recordings and omnidirectional audio from microphone 1 on MAARU's array to BirdNET, using location-specific species lists. Where necessary, we modified these lists to include non-avian sound classes that BirdNET has also been trained to predict, such as 'Engine', 'Human Vocal' and 'Green Frog (*L. clamitans*)'.

In the previous section we only analysed matching predictions in the omnidirectional and beamformed audio. However, this approach omits predictions that occur only in the beamformed or omnidirectional audio. Therefore, here we analysed all predictions of species that were detected at least once in both the omnidirectional and beamformed recordings (henceforth 'common species'), irrespective of whether they occurred at the same timestamp. We did this to test whether beamforming might increase the number of vocalisations detected, for instance, by potentially amplifying bird calls that are too quiet or excessively masked in the omnidirectional recording. We also took this approach to examine whether the omnidirectional recording may comprise a diverse mixture of bird calls from different directions, revealing distinct habitat structure or land use.

To study how the prediction confidence of common species varied between omnidirectional and beamformed audio, we compared species-specific average confidence scores. We took this approach based on more recent advice that 'while there is generally a positive relationship between prediction accuracy and confidence score, the shape of this relationship varies across classes, making it important to evaluate the meaning of BirdNET scores on a class-by-class [i.e., species-specific] basis.' [184, p. 2]. These sets of species-specific mean confidence scores were generally not normally distributed; therefore, we present their median and interquartile range in the Results. In addition, given the above guidance and diverse approaches to setting confidence thresholds across applications and research questions, we compare the number of predictions and median species-specific confidence scores for a range of confidence thresholds (0.25, 0.5, 0.75 and 0.9 as in Section 3.2.3 in the previous chapter).

For three sites (UK Site 2, US Site 2, Malaysia Logged Forest 3) where we beamformed in multiple directions to attempt to distinguish habitat structure and land use, we also

compared the number of predictions with confidence greater than 0.25 for all species identified in each recording (not just common species). We adopted this approach to test whether the different directions correspond to different bird communities.

Table 4.2: Details of beamforming strategy at each recording site for Sub-Project 4. Classes marked 'undesired' refer to types of sound we aimed to minimise by beamforming away from their prevailing source direction. Other classes correspond to target sources of biophony we aimed to beamform towards. We investigate whether BirdNET predicts target and undesired classes more or less frequently in beamformed versus omnidirectional audio from their respective sites. Unless specified, beamforming directions correspond to azimuth angles outlined in Section 4.3.1.

Site Name and Location	**Recorder Orientation**	**Beamforming Direction(s)**	**Beamforming Purpose**	**Target or Undesired Classes**
UK Site 1 (Chickens)	Vertical	Right (to forest patch)	Target sources of biophony	Avian species
UK Site 2 (Pond)	Vertical	Left (to forest patch), Right (to clearing)	Distinguish distinct neighbouring land use	Avian species
UK Site 3 (Road)	Vertical	Left (to forest patch, away from road)	Minimise sources of masking or noise	Engine (undesired)
UK Site 4 (Bees)	Vertical	Right (to clearing and forest patch beyond it, away from road)	Minimise sources of masking or noise	Engine (undesired)
US Site 1 (Geese)	Horizontal	North (to pond with geese)	Target sources of biophony (species-specific)	Canada Goose
US Site 2 (Frog Pond)	Vertical	Left-down ($-135°$ azimuth, to pond with frogs), Up (to canopy)	Target sources of biophony (species-specific); Reveal habitat structure	Green Frog (*L. clamitans*)
Malaysia Oil Palm 1	Vertical	Right (to oil palm, away from anthropophony)	Minimise sources of masking or noise	Human vocal (undesired)
Malaysia Oil Palm 2	Vertical	Left (to oil palm, away from anthropophony)	Minimise sources of masking or noise	Human vocal (undesired)
Malaysia Logged Forest 3	Horizontal	East (to further logged forest), West (to old-growth forest)	Distinguish distinct neighbouring land use	Avian species

For sites where we aimed to use beamforming to target or minimise sounds from specific BirdNET classes (Table 4.2), we also listened to and manually inspected spectrograms of omnidirectional and beamformed audio extracts to help elucidate the results. Specifically, we inspected 60 min of audio from Malaysia Oil Palm 1, 60 min across UK Sites 3 and 4, and the full recordings of US Sites 1 (∼40 min) and 2 (∼3 hr). Spectrograms were primarily inspected in the digital audio workstation software, *Audacity*, and were calculated and plotted for shorter example extracts in *MATLAB* (version 2024a) using the 'pspectrum' function with 0.05 s Kaiser windows, 10% overlap, 0.7 leakage and 1024 Discrete Fourier Transform points.

4.5.2 Results

Increase in Multi-taxa Prediction Counts

Across most sites and confidence thresholds, BirdNET returned more predictions for beamformed audio than omnidirectional audio (Fig. 4.5).[5] Interestingly, the most pronounced exceptions to this trend occurred in sites where we beamformed in two distinct directions. In UK Site 2, the omnidirectional recording returned more predictions than both beamformed recordings for all but the highest confidence threshold. In Malaysia Logged Forest 3, beamforming westwards towards old-growth forest always returned the most predictions, while beamforming eastwards to further logged forest always returned the least, with fewer predictions than the omnidirectional audio. Differences in species-specific average confidence were inconsistent but marginal between beamformed and omnidirectional audio. We also note pronounced differences in the scale of prediction counts across biomes, with thousands of predictions returned for the UK sites over 24 hr, hundreds of predictions for the USA sites with less than a few hours' recording, and only hundreds of predictions over 24 hr in Malaysia – despite Malaysia's rainforests being home to exceptionally high levels of biodiversity.

Specific Target and Undesired Classes

For the five sites where we aimed to beamform away from 'undesired' sounds or towards specific target species (Table 4.2), we found that most sound classes were predicted more

[5]NB. these results do not include any 'engine' or 'human vocal' predictions, which are instead explored in Fig. 4.6.

frequently by BirdNET in the beamformed audio – including those we beamformed away from (Fig. 4.6). The median confidence of target or undesired classes was also the same or marginally lower after beamforming for all sites apart from Malaysia Oil Palm 1, where human vocal predictions for beamformed audio had slightly higher median confidence versus omnidirectional audio. Inspection of spectrograms across all five sites revealed that beamforming suppressed a notable amount of noise present in the omnidirectional recordings (likely a combination of recorder self-noise with some diffuse background environmental noise; Figs. 4.8, 4.9).

For UK Sites 3 and 4, inspecting spectrograms also revealed that a large proportion of engine predictions (the majority in inspected extracts) corresponded to aircraft noise (Fig. 4.9). Some road traffic was however correctly predicted by BirdNET – more so at UK Site 4, as its closest road had heavier and more continual traffic than UK Site 3. While lower frequency traffic noise had audibly similar loudness between the omnidirectional and beamformed audio, higher-frequency sounds such as a car horn were notably quieter after beamforming (Fig. 4.9).

Listening to recordings from Malaysia Oil Palm 1 revealed that most human vocal predictions corresponded to music featuring singing played from a radio[6]. The music was notably louder in the beamformed audio versus the omnidirectional audio. No human vocal predictions were made in the omnidirectional or beamformed audio for site Malaysia Oil Palm 2.

For US Site 1, manual inspection of spectrograms revealed BirdNET correctly identified a high proportion of Canada Goose vocalisations in the omnidirectional and beamformed audio, though several were masked by loud wind noise which was persistent throughout both (owing to the lack of MAARU weatherproofing in this recording). Canada Goose vocalisations were audibly and visually similar between the omnidirectional and beamformed audio. A small number of other bird species' calls (e.g., Red-winged Blackbird and Common Grackle) were notably louder and only predicted in the omnidirectional recording.

Only Green Frog predictions at US Site 2 were more numerous in the omnidirectional audio (29 predictions) than audio beamformed towards the pond they were observed in (left-downwards, i.e., $-135°$ azimuth; 27 predictions). The highest number of Green

[6]Some brief passages contained distinctive radio-like crackle/noise.

Frog predictions was returned when beamforming in an upward direction (bottom right panel, Fig. 4.7, 34 predictions). All Green Frog predictions across the omnidirectional and beamformed audio fell within two ∼7 minute periods in the recordings (13m:27s–20m:18s and 02h:20m:39s–02h:27m:18s). Manual inspection of spectrograms for these periods confirmed that all Green Frog predictions were true positives, but revealed little discernable difference between the omnidirectional and beamformed audio, apart from less noise in the latter (Fig. 4.8). Green Frog vocalisations are short bursts (< 0.5 s) with high intensity across a wide range of frequencies (roughly 0.4–3.5 kHz in our recordings) that resemble the plucking of a string and were audibly almost indistinguishable between the beamformed and omnidirectional audio, though spectrograms visually confirmed they were slightly quieter in the beamformed recordings (Fig. 4.8). However, manual inspection also revealed a low recall rate for Green Frog calls, as only a small proportion of those in the recordings were detected by BirdNET.

Comparing Beamforming Directions

Within sites where we aimed to decompose habitat structure (US Site 2) or distinguish distinct neighbouring land use (UK Site 2, Malaysia Logged Forest 3), differences in the number of BirdNET predictions between beamformed and omnidirectional recordings followed varied patterns for each site (Fig. 4.5). The large increase in predictions for the westward direction (towards old-growth forest) at Malaysia Logged Forest 3 stems from an increased number of Blue-eared Barbet and Sooty-capped Babbler predictions (Fig. 4.7, bottom left panel). The higher number of predictions in the omnidirectional audio for UK Site 2 is driven primarily by European Robin, Stock Dove and Egyptian Goose predictions (Fig. 4.7, top panel). For US Site 2, the slight increase in predictions from the left-downward direction is attributable to increased Red-eyed Vireo predictions (Fig. 4.7, bottom right panel).

The number of predictions for species that were less frequently predicted by BirdNET at UK Site 2 and Malaysia Logged Forest 3 reveal that several were unique to one beamforming direction and/or the omnidirectional audio (Figs. C.2, C.3). However, in most cases these species were only predicted once. A smaller number of species were detected in unique directions at US Site 2, given the shorter recording duration (Fig. C.4).

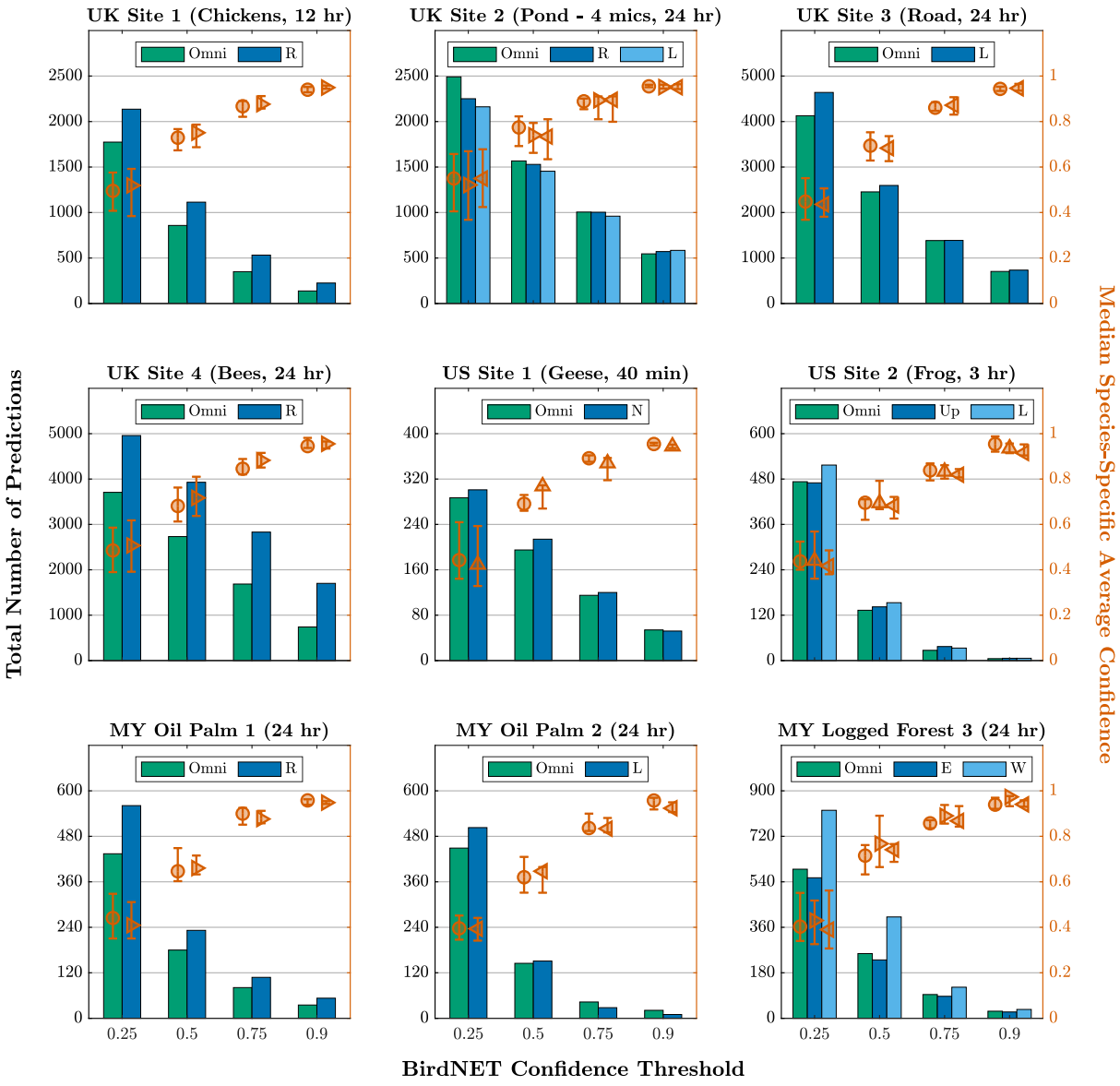


Figure 4.5: Number of BirdNET predictions (bars) and median species-specific average confidence (orange scatter plots) for all recording conditions (omnidirectional and beamformed) across all sites and confidence thresholds from 0.25 to 0.9. Triangles in scatter plot are oriented in the beamforming direction. Orange error bars depict the interquartile range of species-specific averages. Abbreviations for beamforming direction: Omni - Omnidirectional; R - Right; L - Left (Left-downwards for US Site 2); N - North; E - East; W - West.

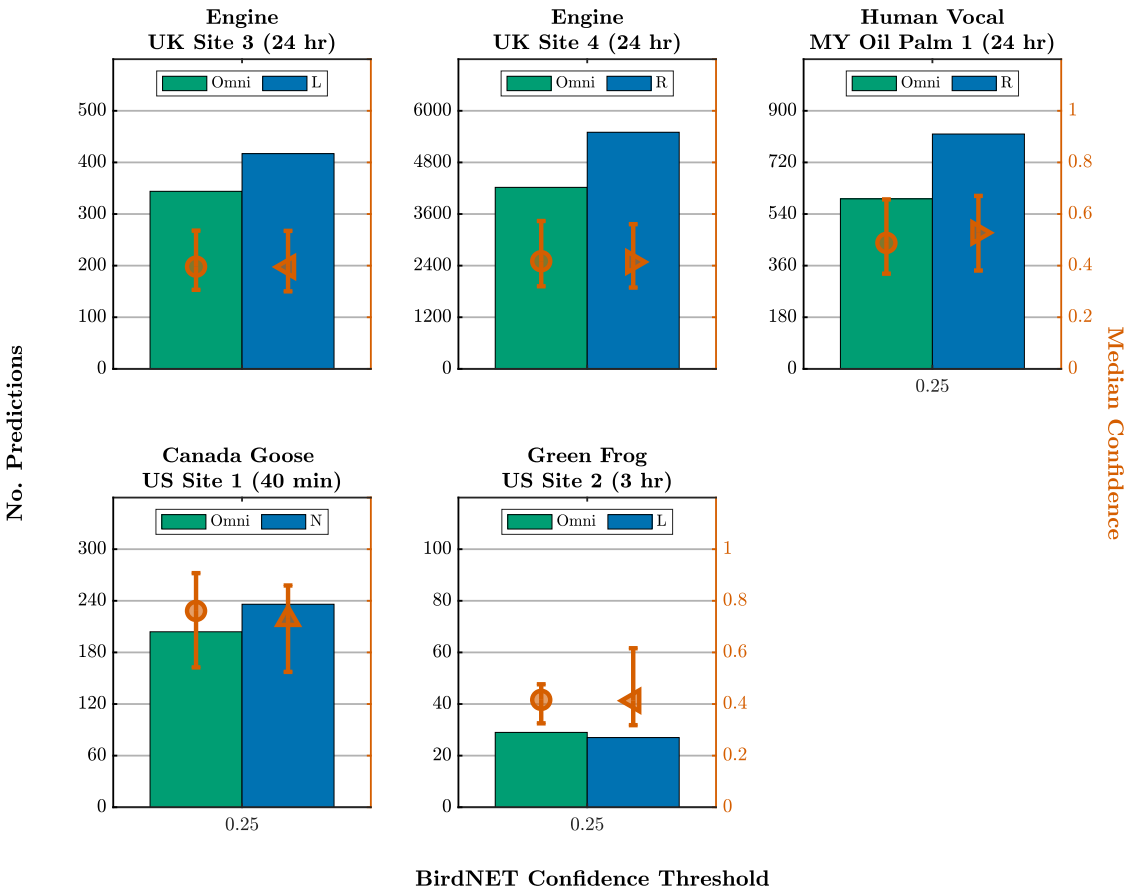


Figure 4.6: Number of BirdNET predictions above confidence 0.25 (bars) and median prediction confidence (orange scatterplots) of individual classes to either be attenuated or amplified via beamforming, for both omnidirectional and beamformed audio. Each panel corresponds to a site where beamforming was intended to attenuate specific unwanted noise (BirdNET's 'Engine' or 'Human Vocal' classes), or amplify specific species ('Canada Goose' and 'Green Frog'). Note that the 'Engine' and 'Human Vocal' predictions are not included in the overall results in Fig. 4.5, and Malaysia Oil Palm 2 is omitted here due to a total lack of 'Human Vocal' predictions. Triangles in scatter plot are oriented in the beamforming direction. Orange error bars depict the interquartile range of confidence scores. Abbreviations for beamforming direction: Omni - Omnidirectional; R - Right; L - Left (Left-downwards for US Site 2); N - North.

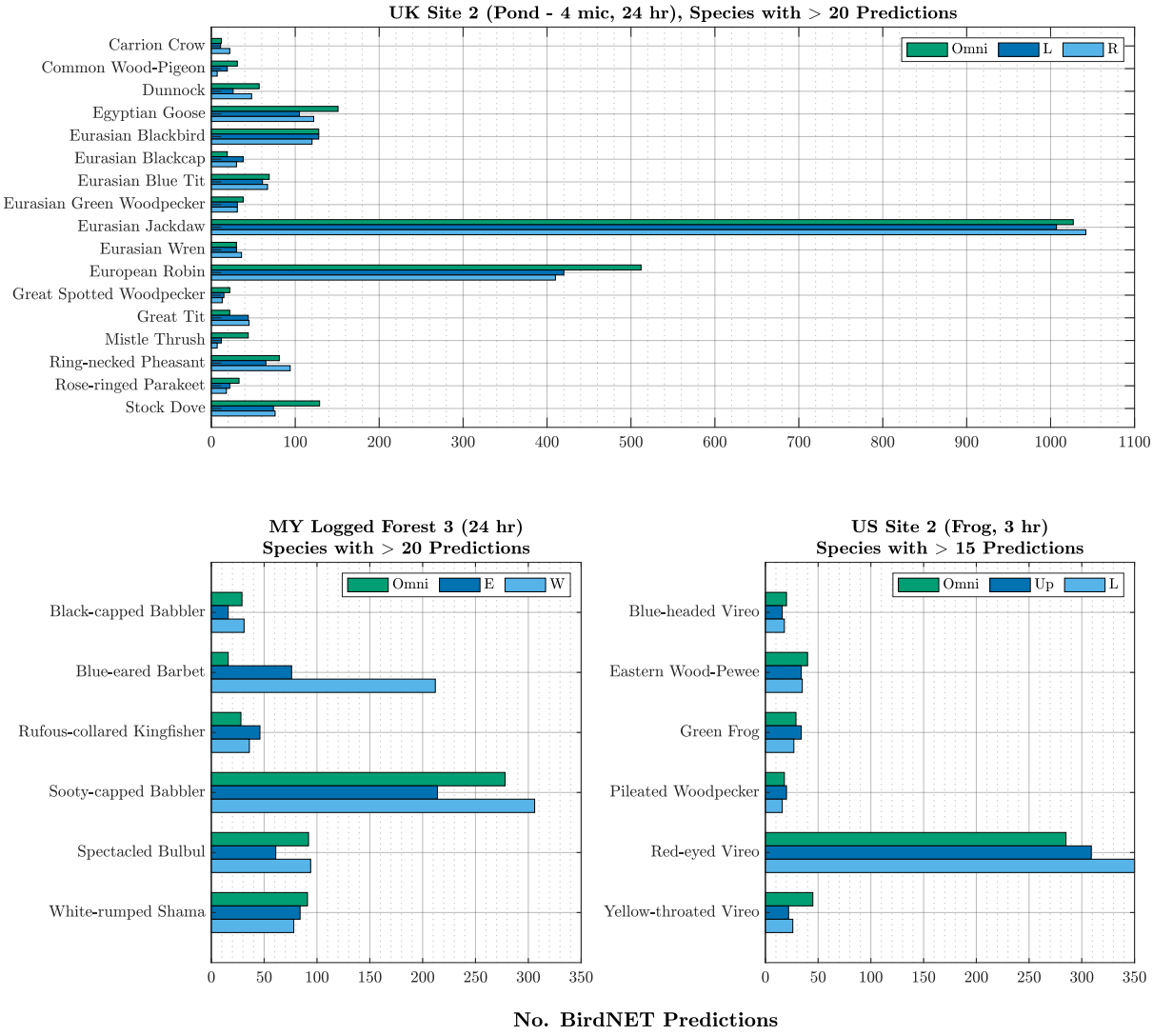


Figure 4.7: Number of BirdNET predictions above confidence 0.25 for most commonly predicted species in omnidirectional and beamformed audio at sites where beamforming was used to attempt to decompose habitat structure (US Site 2) or distinguish neighbouring sites of distinct land use. Abbreviations for beamforming direction: Omni - Omnidirectional; R - Right; L - Left (Left-downwards for US Site 2); E - East; W - West.

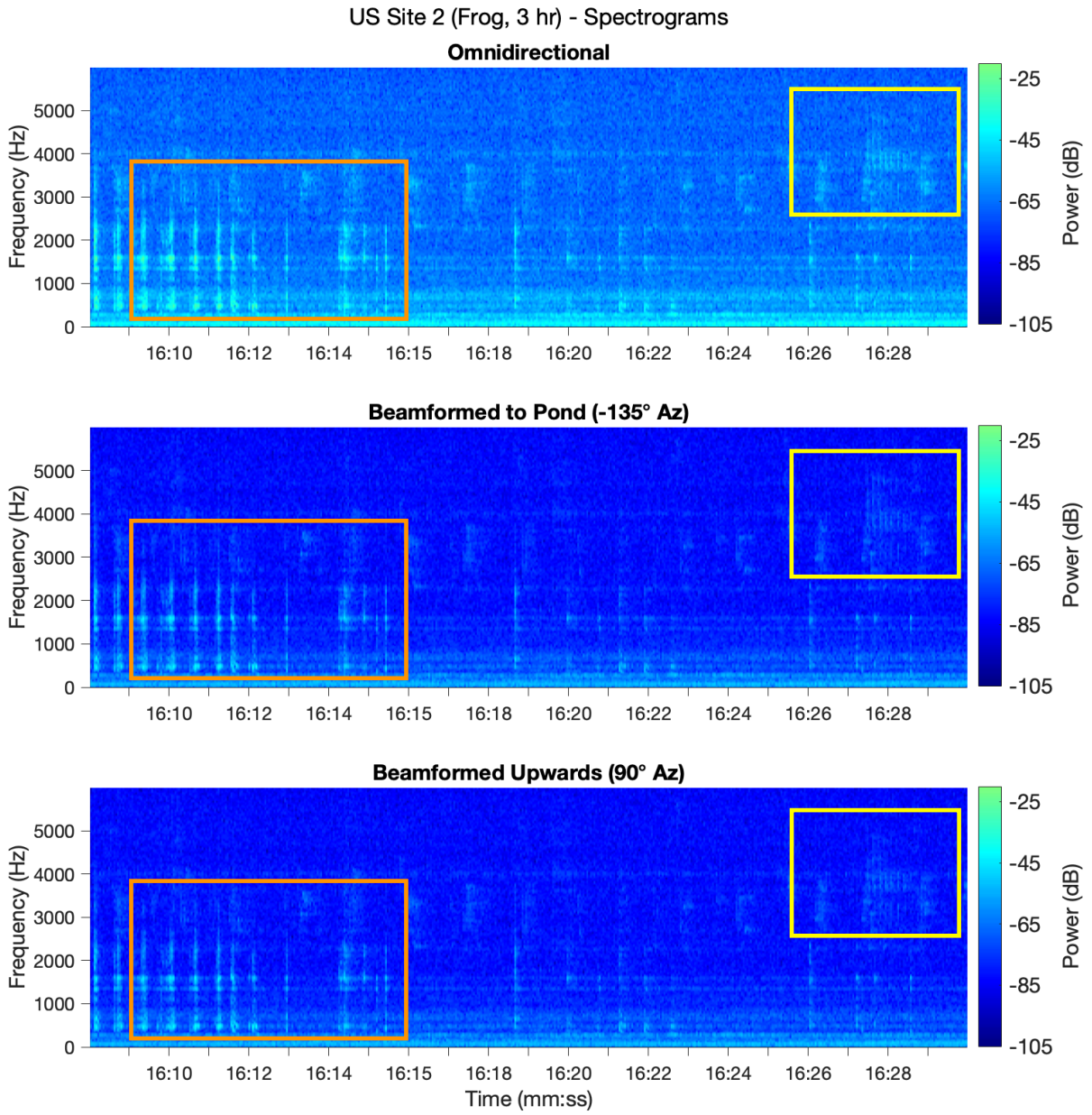


Figure 4.8: Spectrograms showing Green Frog (*L. clamitans*) vocalisations in 22 s extract of omnidirectional and beamformed audio from US Site 2, recorded with MAARU oriented vertically. The orange boxes on the left hand side highlight Green Frog vocalisations that were correctly predicted by BirdNET in all three recordings. (BirdNET makes predictions on 3 s segments of audio, which were set to not overlap in this study, so the 6 s span of these orange boxes corresponds to two consecutive BirdNET predictions.) A number of additional Green Frog vocalisations that can be seen in the spectrograms were missed by BirdNET in all three recordings. The yellow boxes on the right hand side highlight bird calls that were also missed by BirdNET in all three recordings. The Green Frog vocalisations and bird calls are all quieter in the beamformed recordings compared to the omnidirectional recording, though the difference was not audible. However, the notable reduction in noise in the beamformed recordings (darker spectrogram backgrounds) was both visually and audibly perceptible. The x-axis tick labels denote the time into the ∼3 hr recording that the extract corresponds to. The spectrograms are cropped to 0-6 kHz on the y-axes for clearer visualisation. Az: Azimuth.

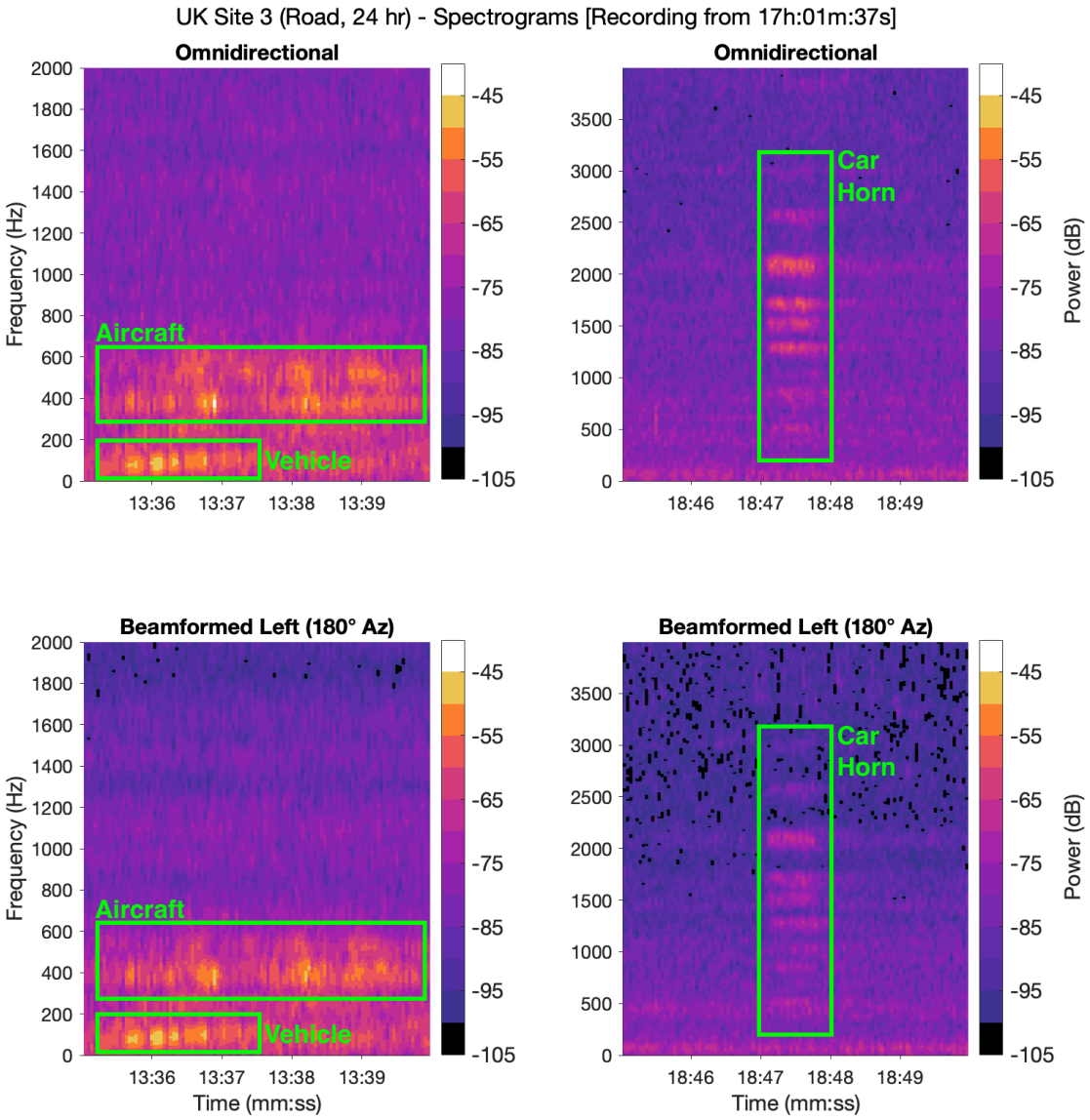


Figure 4.9: Spectrograms showing road traffic and aircraft noise in short extracts of omnidirectional and beamformed audio from UK Site 3, recorded with MAARU oriented vertically. The green boxes and accompanying text denote aircraft noise that mostly sat between 300-600 Hz (left column), a loud accelerating road vehicle with most of its sound power below 200 Hz (left column), and a car horn that spanned a wider range of frequencies (right column). BirdNET returned an 'engine' prediction for the 3 s between 13:36-13:39 (left column) in both the omnidirectional and beamformed audio, though it is ambiguous whether this corresponds to the aircraft or vehicle noise, or both. The aircraft and vehicle noise were audibly indistinguishable between the omnidirectional and beamformed audio, though the spectrograms show slightly less power for the upper aircraft frequencies in the beamformed audio. The car horn (right column) was not predicted by BirdNET but was notably quieter in the beamformed recording, likely because beamforming has greater directivity for higher frequencies. Background noise was audibly and visually lower in the beamformed extracts (darker spectrogram backgrounds). The x-axis tick labels denote the time into the 20 min MAARU recording that the extracts correspond to, which started at 17h:01:37s in the day. The spectrograms are cropped on the y-axes and power values are mapped to a coarse colour gradient for clearer visualisation. Az: Azimuth.

4.5.3 Discussion

Increase in Multi-taxa Prediction Counts

The increases in BirdNET prediction counts for beamformed audio were sometimes relatively small (Fig. 4.5), in part because the omnidirectional audio will have also contained calls from outside the look direction. The beamformed and omnidirectional recordings also likely shared a number of matching predictions. Given that beamforming increased the confidence of matching predictions in Sub-Project 3, the absence of notable increases in confidence here suggests that additional predictions in the beamformed audio may correspond to vocalisations from further afield in the look direction that were too quiet to be detected in the omnidirectional recording. Moreover, increases in beamformed prediction counts were typically largest at the lowest confidence threshold. While these additional beamforming predictions might represent more distant birds, they are likely to be less accurate for many species given their lower confidence scores [184], and should thus be interpreted with caution.

Regional discrepancies in the scale of prediction counts may reflect biases in BirdNET's training data [146] from public wildlife sound repositories, Xeno-Canto and the Macaulay Library. Indeed, BirdNET was first trained on North American and European avian species [167] before expanding to other regions, and existing studies applying BirdNET to global data remarked it predicts fewer species for tropical regions (e.g., [172]).

Specific Target and Undesired Classes

While it was surprising that BirdNET predicted undesired 'engine' and 'human vocal' sound classes more frequently in recordings beamformed away from them (Fig. 4.6), the general increase in predictions for beamformed audio likely stems primarily from the notable decrease in background noise when beamforming (Figs. 4.8, 4.9). This will have led to an improved signal-to-noise ratio for sounds from all directions, including those partially attenuated by beamforming. Moreover, road vehicle and aircraft engine noise at UK Sites 3 and 4 predominantly contained frequencies below 1 kHz (in line with prior literature for road traffic [284]), for which MAARU's directivity pattern is not particularly directional after steering (Fig. 4.2). This explains the similar loudness of these sounds between omnidirectional and beamformed recordings. However, the attenuation of higher

pitched sounds associated with road vehicles and aircraft (e.g., upper aircraft frequencies and car horn in Fig. 4.9) indicates that beamforming was effectively steering away from these sources for higher frequencies. In Malaysia Oil Palm 1, the increased number and median confidence of human vocal predictions when beamforming away from human settlements and towards oil palm plantations appeared to stem from human activity in the plantations themselves. Specifically, human vocal predictions corresponded to music containing vocals from a radio that was louder when beamforming to the plantations.

Though small, the increase in Canada Goose predictions for beamformed audio was expected. The increase was most likely slight because many geese were in close proximity (up to tens of meters) to the recorder, and thus equally detectable in the omnidirectional and beamformed audio, while wind noise that remained prevalent in both masked a similar proportion of calls. The decrease in median beamforming confidence again suggests beamforming captured more distant vocalisations in the look direction. Both this and the attenuation of bird calls from other directions[7] reinforce that beamforming is effective at steering MAARU's recordings for higher frequencies.

The results for Green Frog predictions initially appeared perplexing, with a higher number of predictions in the omnidirectional audio than audio beamformed towards the pond they were observed in, and the highest for audio beamformed upwards. Green Frogs are not known to climb, but do forage in land with high vegetation density and leaf litter near the bodies of water they inhabit [285, 286]. A plausible explanation could thus have been that vocally-active frogs foraging around the pond were picked up in the omnidirectional audio and that beamforming upwards captured reflections of all calls from the canopy [182]. However, these prediction count discrepancies should be interpreted with caution given that Green Frog calls were present and spectrally similar across the omnidirectional and beamformed recordings and that differences between the BirdNET prediction counts were small relative to the number of missed calls. In reality, BirdNET's Green Frog predictions were comparable – with similarly poor recall – across the US Site 2 recordings. The low recall for Green Frog predictions may also be reflective of BirdNET's limited training on anurans and other non-avian species [167][8], while the small differences in predictions may have arisen from the subtle, inaudible differences between

[7]For instance, the attenuated Red-Winged Blackbird and Common Grackle were likely at nearby feeders and trees west of MAARU, relative to its coordinate system (Section 4.3.1).

[8]Though the model has been updated to include more non-avian classes since its initial training – see https://birdnet-team.github.io/BirdNET-Analyzer/models.html

the omnidirectional and beamformed recordings.

Comparing Beamforming Directions

Using beamforming to distinguish neighbouring land use appears to have been more successful in Malaysia Logged Forest 3 than UK Site 2, given the former's more distinct prediction counts for different beamforming directions across confidence thresholds and for various individual species (Fig. 4.5, 4.7). This is perhaps unsurprising, given Malaysia Logged Forest 3 is surrounded by a much vaster expanse of natural habitat comprised of two larger and more distinct types of land use – old-growth and logged forest – than UK Site 2. In general, all UK sites were much more heterogeneous over a smaller spatial scale than the Malaysia sites, comprising localised patches of different vegetation and land-use types. Moreover, Silwood Park is surrounded by roads – including a major 'A' road to its south – with traffic that would likely mask biophony from further afield.

Nonetheless, it is promising to see that beamforming at Malaysia Logged Forest 3 reflected patterns reported elsewhere [287] of higher biodiversity (meaning here both more BirdNET predictions, per Fig. 4.7, and more species detected, per Fig. C.3) from the direction of old-growth forest compared to the direction of further logged forest. As noted for other sites, the higher species counts for omnidirectional over logged-forest-oriented audio likely arise from the omnidirectional recording containing sounds from both directions. Further, the consistently lower median species-specific average confidence for the old-growth beamforming direction (Fig. 4.7) may also be indicative of quieter vocalisations from further within intact forest being captured.

The lack of such distinct directional patterns for UK Site 2 (Fig. 4.7) most likely stems from the use of four microphones for beamforming (configured as in Fig. 4.2), caused by the failure of two microphones on this particular MAARU. As shown in Fig. 4.2, using these four microphones introduces much more pronounced side lobes with time-delay beamforming, especially for frequencies above 4 kHz, which are impacted by spatial aliasing (Section 4.3.1). As such, many directions around the array – corresponding to the main lobe and side lobes in Fig. 4.2 – will have been captured when beamforming to either look direction. The omnidirectional recording nonetheless returned more predictions than beamforming with four microphones (Fig. 4.5, 4.7), perhaps because callers were present at directions still attenuated in the beamforming power pattern. This is corroborated by

BirdNET returning more predictions on audio beamformed with the three microphone subset shown in Fig. 4.2, which is even less directional. In particular, beamforming with these three microphones in a 0° azimuth look direction produces side lobes with greater power at 30° (8 kHz), ∼40° (6 kHz), ∼60° (4 kHz) and 90° (8 kHz) azimuth. When beamforming rightwards (180° azimuth), this pattern will be mirrored and correspond roughly to directions facing the forest patch canopy. This may explain why beamforming rightwards (180° azimuth) with three microphones returned the most BirdNET predictions at UK Site 2, both overall across confidence thresholds (Fig. C.1) and for a number of specific species (Fig. C.5).

4.6 Outlook and Conclusions

In this chapter, we have taken steps towards the efficient, long-term autonomous spatial monitoring of biodiversity using multichannel acoustic recording and beamforming to decompose complex soundscapes. We first iteratively redesigned and field tested the weatherproofing of MAARU, to improve its robustness to extended environmental exposure. Our final iteration balances off-the-shelf and 3D-printed components, is low-cost, provides easier access to internal components, and is more resistant to animal and insect interference. However, our final deployments encountered other failure modes involving water ingress and under-powering. We thus shared recommendations such as sheltering the devices or deploying them upside down and using 100 W panels for 24/7 monitoring, ensuring future deployments can gather data for months or longer.

We then evaluated the performance of simple, directionally-static beamforming with MAARU in theory and in practice, relating the impact of beamforming on bird call classification to array layout and geometry. We found that beamforming can boost biophonic signals, typically returning more confident and/or numerous BirdNET predictions than omnidirectional recordings across diverse taxa and regions from tropical to temperate forests. This likely stems from beamformed audio featuring a higher signal-to-noise ratio, more closely resembling the focal recordings BirdNET was trained on, and potentially capturing more distant vocalisations while attenuating higher frequency sounds from some non-target directions. We have also shown potential for distinguishing habitat structure and differences in bird community composition between neighbouring sites, with par-

ticularly promising results near the edge of old-growth and logged forest in Malaysian Borneo.

However, our more surprising findings point to imperfections in both MAARU and directionally-static time-delay beamforming as tools for spatial filtering. These imperfections limit the extent to which soundscapes can be cleanly decomposed into separate spatial segments. As such, beamforming's main advantages at present are its enhanced signal-to-noise ratio and focal-like nature. Immediate next steps should further probe our results by comparing more spectra of omnidirectional and beamformed audio across our recordings, and using acoustic indices and deep learning model embeddings to explore the extent to which beamforming in different directions is capturing distinct subsets of soundscapes. It would be particularly interesting, for instance, to investigate whether model embeddings from beamformed audio form sub-clusters of the omnidirectional recordings' embeddings.

Spatially segmenting soundscapes is nonetheless possible, but requires careful consideration of the landscape, recorder properties and deployment strategy, including how microphone array configuration shapes beamforming directivity. Future work should explore directionally adaptive beamforming techniques as well as other ways that directionally static beamforming could be enhanced, for instance with optimum beamforming strategies or integration with monophonic source separation techniques. The latter could reduce data loads via on-device directionally static beamforming that captures sounds from a few salient directions, while still allowing for more fine-grained source identification in each look direction via monophonic source separation on the server side. Future innovations are also required in fundamental aspects of hardware and software that currently limit scalability across the full recording and processing pipeline, such as more efficient solar panels and microprocessors, cheaper arrays with more microphones, and more lightweight techniques for optimum and/or directionally adaptive beamforming. Combining such innovations could allow the spatial composition of soundscapes to become a key focus of the automated passive acoustic monitoring of ecosystems.

Chapter 5

Multimodal assessment of tropical rainforest resilience loss

Summary

1. Climate change and human activity are eroding tropical rainforest resilience, risking tipping points of mass forest dieback towards savannah-like or treeless states. Here, we explore satellite and acoustic data for early warning signals (EWSs) of rainforest tipping points.

2. We first extend a close analysis of tipping point EWSs in satellite-sensed vegetation from the Amazon to Borneo, revealing resilience loss manifest as critical slowing down of recovery from perturbation. Specifically, we examine spatiotemporal patterns in rising lag-1 autocorrelation (AC(1)) and variance to assess whether, when and where vegetation resilience loss accelerated in Borneo, and how these patterns relate to climate, human activity and animal biodiversity. We find Bornean rainforests lost resilience over 1991–2016, with a marked acceleration from a breakpoint in ∼2004. Resilience loss accelerated earlier in subregions with higher human footprint, and was more severe in subregions with lower precipitation and higher human footprint before the breakpoint. Post-breakpoint, resilience loss was stronger in wetter and less biodiverse areas. We also uncover the apparent reallocation of resilience loss post-breakpoint to areas that lost less resilience pre-breakpoint. Climate and bird and mammal

richness appear to have contributed to this pattern most.

3. We then explore calculating EWSs in acoustic descriptors of sound-producing biodiversity as a potential route to gain further insight into the relationship between animal biodiversity and vegetation resilience. We examine lag-1 autocorrelation and variance in acoustic indices extracted from recordings over a two-year period in the Amazon and along a land-use gradient in Borneo. We find that larger scale data is needed to conclusively determine whether critical slowing down manifests in soundscapes. However, we see some potential in using EWSs to capture changes to fluctuations in acoustic data at various temporal scales. Further, we find nascent evidence of slower vocal biodiversity dynamics in logged versus old-growth forest in Borneo.

4. Overall, our results provide enhanced evidence of Bornean rainforest resilience loss and its potential drivers, underscoring the urgent need for ecological conservation and climate change mitigation in Borneo and worldwide.

5.1 Introduction

Tropical rainforests are essential for global biodiversity and ecosystem services, supporting over two-thirds of the world's terrestrial biodiversity and providing critical functions in climate regulation, water cycling, and carbon sequestration [288–292]. However, over the past few decades, tropical regions have experienced some of the most severe forest loss [293,294]. Borneo has lost considerable forest to drought, fire, deforestation, and other pressures [293–297], and there is mounting evidence that climate and land-use change are driving tropical forests towards abrupt fold bifurcation tipping points (Section 2.1.1; Fig. 2.1b,d,f) involving mass forest dieback towards alternative savannah-like or treeless states [62–64, 84, 113]. The threat of such tipping points is alarming as they can be difficult to reverse and their approach cannot be inferred from gradual changes to the mean state of the system alone [74, 83, 190]. In this chapter, we build on existing work using generic EWSs to track systemic declines in tropical rainforest resilience by adopting a multimodal approach. We first analyse spatiotemporal patterns and possible drivers

of vegetation resilience loss in Borneo using satellite data. We then explore acoustic recordings from the Amazon and Borneo for tipping point EWSs to probe the resilience of soniferous rainforest biodiversity at various temporal scales.

We view tropical forest ecosystems as complex dynamical systems with savannah and treeless states as alternative attractors, and interpret resilience loss as a declining capacity to withstand disturbance, exhibited as the critical slowing down of recovery from perturbation, as outlined in Chapter 2. We focus on generic statistical EWSs of tipping points, namely variance and autocorrelation (specifically, AC(1)), that detect signatures of critical slowing down in a system's fluctuations about its long-term mean. These archetypal EWSs have been used to track the resilience of forests and the biosphere at large in decades of high-resolution satellite vegetation data (e.g., biomass, water content [84, 113]). While this work has revealed worrying global trends in forest resilience loss, more detailed analyses, such as identifying potential drivers of resilience loss [62] and examining alternative stable states empirically [63, 64], have focussed primarily on the Amazon rainforest. Using over 20 years' satellite vegetation data, Boulton et al. [62] revealed a marked loss of Amazon rainforest resilience from the early 2000s, with faster loss (i.e., more consistent rises in AC(1) and variance) in regions with lower mean annual precipitation and greater proximity to human activity.

Existing studies have mostly focussed on relating observations of resilience loss to climate and human pressures, given that declining precipitation, deforestation and fire are understood to be key drivers of self-propelling dieback in tropical rainforests [63, 64] (see Section 1.1.3). However, internal ecological drivers, particularly biodiversity, have received less attention. The relationship between organisms and their environment is a cornerstone of ecology and the impact of biodiversity on ecosystem stability has been researched for decades (e.g., [218, 298, 299]). Indeed, one of the first examples of a fold bifurcation in ecology was in a model of a grazing system [98, 99]. However, most studies on the role of biodiversity for ecosystem resilience have focussed on community-level processes and interactions [300–302] and the impact of functional diversity in mitigating tipping points to alternative states within various animal, fish and plant communities [70]. Relatively little work has empirically studied whether and how animal and insect biodiversity relate to nonlinear dynamical system models of rainforest vegetation. Prominent researchers in this space have, however, called for further research assessing the reciprocal interaction

between the dynamics of population traits and large-scale ecosystem regime shifts [303]. Other work has also explored whether EWSs of population collapse emerge for various taxa along a gradient of tropical forest loss [111]. Importantly, Flores and Holmgren [304] found that Amazonian floodplains' low resilience [305] and inability to recover from multiple fires may partly be driven by burnt floodplains' low tree cover failing to attract frugivores such as birds and mammals, which limits seed arrival and thus prevents forest regrowth. Given animal biodiversity's fundamental role in ecosystem function, we posit that even if animal populations do not exhibit fold bifurcations as rainforests tip into savannah, they may still experience resilience loss including critical slowing down as vegetation nears its threshold for mass dieback.

As discussed in prior chapters, acoustics can be particularly effective for monitoring biodiversity and ecosystem health at many levels [13, 14]. In the context of resilience, acoustic monitoring has primarily been used to study the recovery of soniferous species and vocal communities from large-scale disturbances such as fires and hurricanes [26, 132, 184]. This has allowed for communities' resistance and recovery to large perturbations to be directly estimated (e.g., [132]). Furthermore, we know that anthropogenic pressures, natural habitat destruction and associated biodiversity declines are causing natural soundscapes to fall silent and change in their composition and temporal structure [26, 27]. However, soundscapes have yet to be analysed for long-term signs of critical slowing down, which could reveal a progressive loss of resilience in soniferous populations' recovery from continual small perturbations and stochastic noise, rather than their response to isolated large disturbances. Changes to soundscapes' content and resilience-related statistical properties (e.g., variance and AC(1)) across different temporal scales may thus reveal how various subsystems and communities that make up complex ecosystems are losing resilience. Indeed, prior research using acoustics to measure ecological community responses to major disturbance (a hurricane) called for soundscape variability to be studied as an indicator of ecosystem disturbance and potential warning of regime shifts [132].

Here, we extend the methods used by Boulton et al. [62, 306] to quantify changes in Amazon rainforest resilience to Borneo. We measure the AC(1) and variance of satellite-sensed vegetation data for Borneo – incorporating recent advice on aggregating satellite data with sensor shifts [190] – and use breakpoint regression to provide additional quantitative metrics for the timing of changes in these EWSs' trends. We also consider a broader

set of regional factors that may be driving rainforest resilience loss – precipitation, temperature, human activity and animal biodiversity – and compare their associations with resilience change before versus after the breakpoint. With this satellite-based assessment of resilience as our backbone, we then explore whether rising AC(1) and variance manifest at different temporal scales in acoustic biodiversity indices from soundscape recordings of the Amazon and Borneo. Specifically, we look for the critical slowing down of acoustic indices' diel fluctuations (i.e., fluctuations over 24 hr) in the Amazon and their intra-diurnal fluctuations (i.e., short fluctuations within the daytime) in Borneo. This serves as an initial exploration into applying the framework for acoustically monitoring ecosystem tipping points we outlined in Chapter 2. We evaluate our findings in relation to this framework in the Discussion (Section 5.4.3). Through our analysis, we provide more detailed insight into patterns of resilience loss in Bornean rainforest and the factors associated with it. We also expand the early-warning toolkit for critical transitions, to: (i) explicitly quantify the timing and magnitude of resilience loss; and, (ii) potentially complement satellite data with acoustic recordings to anticipate ecosystem tipping points and better understand their drivers, including biodiversity.

5.2 Materials and Methods

5.2.1 Satellite Data Analysis

Datasets and Pre-processing

For our satellite-based analyses, we defined the geographical extent of Borneo (7°N to 4°S, 108°E to 119°E) by integrating shapefiles of Malaysia, Brunei, and Indonesia from the GADM database version 4.1[1]. To identify the intact rainforest of Borneo as our study area, we applied filters to the International Geosphere-Biosphere Programme (IGBP) Moderate Resolution Imaging Spectroradiometer (MODIS) land cover dataset MCD12C1 v061 [307]. We selected grid cells where evergreen broadleaf forest cover was $\geq 80\%$ in 2001 and excluded any cells containing human land use (built-up, croplands, or vegetation mosaics $> 0\%$; [62]).

To evaluate the resilience of intact Bornean rainforest, we used Vegetation Optical Depth (VOD) data from the Vegetation Optical Depth Climate Archive (VODCA) [308].

[1] https://gadm.org/data.html, accessed 17 November 2023.

Compared to optical remote sensing vegetation products, such as the Normalized Difference Vegetation Index (NDVI), passive microwave-derived VOD offers high sensitivity to high biomass levels, and the capability to perform under cloud cover, making it particularly suitable for tropical forest analysis [309–311]. VOD products with shorter wavelengths are more sensitive to leaf moisture content and better reflect changes in vegetation resilience, but are less influenced by stem biomass and soil moisture [308,312,313]. The Ku-band of VODCA was selected because it is the shortest wavelength band of the dataset, while providing the longest time span of short wavelength products by combining multiple sensors [308, 312].

To explore factors that may affect resilience change, we considered three perspectives: climate, human activity, and biodiversity. For climate, we focussed on precipitation and temperature, sourced from the monthly average precipitation data of the Climate Hazards Center InfraRed Precipitation with Station data (CHIRPS) v2.0 dataset [314] and the fifth generation of European ReAnalysis (ERA5) Land monthly average 2 m air temperature dataset [315], respectively. Impacts of human activity were assessed using human footprint data from the database described by Mu et al. [316], which provides a global measure of human pressure on the environment based on eight variables, including human population density, proximity to roads, and levels of built environment and nighttime light. Areas with values below 1 are interpreted to be wilderness, those between 1 and 4 are deemed intact, and those over 4 are deemed modified (values may be as high as 50). To explore the relationship between animal biodiversity and rainforest resilience, we utilised animal species richness data from the International Union for Conservation of Nature (IUCN) Red List version 2022-2 [317], including four taxa (amphibians, birds, mammals, and reptiles).

Preliminary exploration of the data highlighted significant fluctuations in the first three years of VODCA data, which were also reported in [62]. Therefore, we also restricted our analysis to 1991–2016. The land cover dataset was only available from 2001 and was used up to 2016 to match the study period. We adopted a $0.25° \times 0.25°$ spatial resolution, matching that of the VOD data (the coarsest dataset we examined) and unified the data under the WGS 1984 geographic reference system. We then aggregated data with different collection frequencies to the required analysis scale and cropped all data into the study area using the package terra [318] from *R* (v. 4.4.1; [319]).

Generation of Resilience Indicators for Satellite Data

We used the VOD time series to estimate the resilience indicators, AC(1) and variance. We first applied the seasonal and trend decomposition using Locally estimated scatterplot smoothing (Loess) to remove the long-term trends and seasonal fluctuations from the VOD data [320], in order to obtain a stationary time series of fluctuations in the system state, caused by small perturbations and noise [189, 321]. We used the 'stl()' function in *R* (v.4.5.0; [319]) to decompose the time series in each grid cell (using the 'periodic' option for seasonal windows), retaining the residuals for subsequent analysis [62].

AC(1) was calculated by fitting an order 1 autoregressive model [107]:

$$x_{t+1} = \alpha_1 x_t + \varepsilon_t \tag{5.1}$$

fit using ordinary least squares with Gaussian random error ε_t. This formula does not include an intercept because the model's analysis was performed on detrended residuals with a mean of zero.

We calculated resilience indicators over a 5-yr (60-month) monthly sliding window, creating a new time series from July 1993 to July 2014. Importantly, to calculate overall resilience indicators for the entirety of Borneo, we first spatially averaged the VOD time series and then computed resilience indicators on the mean VOD time series. This order of operations has proven more effective in reducing bias, especially that which arises from mixing sensors, compared to averaging resilience indicators computed for each grid cell [190]. To explore the spatial variation of resilience loss across Borneo, we calculated the resilience indicators for each grid cell following the same sliding window approach.

Previous work has shown that AC(1) is a more robust indicator when analysing resilience [189, 322], especially as variance is more susceptible to noise in ecosystem data and changes in environmental perturbations, hence we focus primarily on AC(1) in our analyses. However, Smith et al. [190] recently highlighted that combining data from different sensors, as with VODCA, may also contribute to a rise in autocorrelation. They also note that combined sensor data may bias AC(1) and variance toward negative correlation with one another, whereas actual resilience loss should yield increases in both AC(1) and variance [62, 190]. Therefore, we mainly employ variance as a supplementary metric to validate that the increases in AC(1) are likely due to resilience loss. To further

examine the potential influence of satellite sensor transitions on the VODCA products, we incorporated the timing of these shifts into the mean AC(1) series and compared these with the breakpoint (Fig. D.1).

Metrics of Resilience for Satellite Data

A Bayesian changepoint detection algorithm was used to reveal when trends in resilience changed (i.e., the **breakpoint**), employing the package 'mcp' (v.0.3.4; [323]). We used **Kendall's τ** non-parametric rank correlation coefficient to assess the consistency of changes in the AC(1) time series for the full, pre- and post-breakpoint study periods. Kendall's τ returns values of 1 for a consistently increasing series, -1 for a consistently decreasing series, and 0 when there is no overall trend. To specifically examine the magnitude of the rate of resilience loss around the breakpoint, we also extracted the posterior mean slope values (henceforth, '**slope**') of the linear segments before and after the breakpoint. Within each segment, positive slope indicates resilience loss, negative slope indicates resilience gain and larger absolute values denote a greater magnitude of resilience change.

The p-value directly derived from the Kendall's τ coefficient may be inaccurate due to its sensitivity to the autocorrelation characteristic inherent in time series data. Therefore, following prior work, we employed a phase randomization procedure to generate surrogate data to determine the significance of Kendall's τ [62, 107, 324]. Specifically, we performed a Fourier transform on the AC(1) time series, randomly permuted the phases, and then applied an inverse Fourier transform, thereby maintaining the autocorrelation and variance in the surrogate series [62, 325]. We performed this 100,000 times to create a null model for assessing the significance of Kendall's τ. The p-value was determined by calculating the proportion of surrogates that exhibit a higher Kendall's τ value than that observed in the actual series.

Correlates of Vegetation Resilience

To focus on changes in resilience and standardise the four factors we studied for associations to Bornean rainforest resilience (mean annual precipitation, mean annual temperature, human footprint and vertebrate species richness), we treated each factor as constant. Factors with time-series data, such as precipitation and temperature, were thus

averaged over the study period for subsequent analysis, allowing us to focus solely on the spatial patterns in the factors and resilience change. We also assessed potential spatial correlations between the factors by calculating their pairwise Spearman's rank correlation coefficients (ρ) (selected given several factors had non-normal distributions and could have nonlinear relationships).

Building on previous work [62, 113], we explored the impact of these factors on resilience by ranking the grid cells across our study region according to their value for each factor, and calculating the VOD AC(1) time series for sliding bands of 100 grid cells that advanced through the ranked cells one cell at a time. The VOD AC(1) time series for each band of 100 cells was calculated using the same order of operations as for the entirety of Borneo (i.e., spatially averaging the VOD signals first, then calculating AC(1)). For each band, we calculated our three resilience metrics (breakpoint, Kendall's τ and slope) on the corresponding AC(1) time series. This procedure reduces noise from other factors and effectively reveals how our resilience metrics varied along spatial gradients of each factor. Finally, we used Kendall's τ to assess the tendency of correlations between the resilience metrics and gradients of factors.

Robustness Tests

To test the robustness of the results, we estimated the AC(1) trend when varying the evergreen broadleaf forest cover threshold to $\geq 60\%$ and $\geq 90\%$ (Figs. D.3, D.4). We also varied the sliding window length used to calculate AC(1) to 30, 90, and 120 months (Fig. D.5). For exploring the relationship between factors and resilience loss patterns, we also used sliding bands of 50 grid cells (Figs. D.10, D.11).

5.2.2 Acoustic Data Analysis

To explore EWSs of rainforest tipping points in the acoustics domain, we examined various temporal patterns in acoustic indices calculated on months of recordings collected over a two-year period in the Amazon and one week of recordings along a land-use gradient in Malaysian Borneo. We focussed on two prevalent acoustic indices typically used as proxies of biodiversity (in particular of avian diversity): the acoustic complexity index and the bioacoustic index (see Section 2.2.3 for technical background on each index).

Recording Locations and Schedule

Recordings in the Amazon were gathered at the middle of three permanent 300 m transects maintained by the Sustainable Amazon network [326] in *terra firme* forests near Santarém, Belterra and Mojuí dos Campos (3.046°S, 54.947°W), Brazil. Soundscapes were recorded semi-autonomously over a period of two years from January 2018 to December 2019 using *Frontier Labs* bioacoustic recording units. The recorders were set to record one in every 10 mins, 24 hr/day, at 44.1 kHz sampling rate and 16-bit uncompressed WAV (waveform audio format) recording. Around a quarter of files were lost due to technical failures. After recording, the audio dataset was then cleaned to remove recordings with heavy rain or other problematic sounds (e.g., issues with the recording hardware) prior to calculating acoustic indices[2]. This left around two thirds of the full recording period available for analysis.

In Borneo, recordings were captured autonomously during the MAARU deployments described in Section 4.2.1. We identified a week-long period, 13-19 September 2024, during which there was near-simultaneous recording in one of each of our land-use types (old-growth forest, logged forest and oil palm plantations). Recordings in the replicate sites had little to no overlap with this period owing to staggered deployments and recorder failure (Section 4.2.1), hence they were not studied here. In old-growth forest, a smaller solar panel was used, resulting in the recorder losing power overnight. Old-growth recordings were thus mostly limited to around 07:30 to 20:30, resulting in 52% coverage of the recording period. In logged forest, recordings were mostly captured for 24 hr/day, but some files were lost due to a daily scheduled reboot at 02:00 interrupting recording. On one day (17 September 2024), the logged forest recorder failed to record after rebooting[3] and only streamed five 20 min files, resulting in around 85% coverage of the recording period in logged forest overall. The oil palm units recorded 24 hr/day, losing a small number of files on reboot with 98% coverage of the week.

In order to capture the full soundscape at each site, we extracted the omnidirectional recording from MAARU's upper left microphone. We did not perform any data cleaning, given the short duration of the recording period and lack of data due to powering and connectivity issues on certain days.

[2]Further details on the recording sites, setup and data cleaning are provided in [181].

[3]Either due to insufficient power or a failure to re-establish an internet connection after rebooting and a lack of onboard storage.

Calculation of Early Warning Signals in Acoustic Indices

We opted to consider the acoustic complexity index and bioacoustic index as they have previously been studied on these Amazon recordings and further recordings from the same region[4], revealing various correlations to animal [144] and insect [181] biodiversity. Both studies noted stronger correlations between the bioacoustic index and different aspects of biodiversity, whereas the acoustic complexity index sometimes showed negative correlations to avian biodiversity [144], in contrast to findings from prior literature [156–158]. However, these correlations were based on taxon-specific frequency ranges (3-10 kHz for insects in [181] and 0.3-4 kHz for avian biodiversity [144]). Here, we examine the values of the indices for a wider range of frequencies: 0.3-10 kHz for the Amazon and 0.3-8 kHz (limited by the Nyquist frequency for MAARU's 16 kHz sampling rate), to allow for a wide range of taxa to be included [26] while reducing potential recorder self-noise below 300 Hz [144].

Procedure for Amazon Indices For the Amazon data, the acoustic complexity index and bioacoustic index were calculated on the cleaned 1 min sound files using the *soundecology* [256] package in *R* [319] (i.e., acoustic index window: 1 min; resolution: every 10 min, except when there were gaps in recording). To verify whether the indices followed expected diel and seasonal trends, we calculated their average diel values at Brazil Site 1 for the full recording period at and for each month of the year. We then selected periods of two weeks or more that contained gaps in data of no more than seven consecutive days, in order to calculate EWSs on relatively continuous time series of reasonable duration.

Next, we averaged the indices to hourly resolution and imputed any gaps with the mean value(s) for the missing hour(s) in the corresponding month. Recalling that a system must be sampled more frequently than its return time to perturbation in order to detect critical slowing down, we expect this hourly resolution to capture any changes to the diel fluctuations of the indices. These changes may in turn be indicative of resilience loss amongst soniferous biodiversity, especially given evidence that human disturbance including logging and oil palm plantations are weakening temporal soundscape patterns in Amazonian [26] and Bornean [27] rainforests. This hourly sampling rate could in theory also capture changes in slower soundscape fluctuations that might reveal shifts in vegeta-

[4]Captured in primary, logged, burned and secondary forest using the same recording equipment (*Frontier Labs* bioacoustics recording units).

tion biomass, through the impact of vegetation on acoustic reflections (reverberation) at each site [182], and because of the relationship between vegetation and animal and insect biodiversity (e.g., through fruiting). This, however, will partly depend on the nature of subsequent pre-processing, including detrending and the sliding window size [186].

We implemented Locally weighted scatterplot smoothing (Lowess) detrending[5] in the Python package *ewstools* [327]. Most long, near-continuous sections of data across the two-year period were roughly several months in length, so after some initial experimentation we used a Lowess span of two weeks (336 data points for hourly data). This preserved the diel fluctuations in the data while reducing trends on the order of multiple weeks. Finally, we calculated lag-1 autocorrelation[6] and variance on two-week sliding windows advancing hourly for each imputed section of data, and computed their Kendall's τ values, using the *ewstools* [327] Python package.

Procedure for Borneo Indices Given the Borneo recordings were comparatively short-term and captured along a land-use gradient, we aimed to apply the same tipping point EWSs to the acoustic indices to reveal any changes to fluctuations *within* diurnal soundscape patterns in increasingly human-modified environments.

We used the *scikit-maad* (version 1.5.1) Python package to calculate the acoustic complexity index and bioacoustic index for every 10 min period in the recordings from each site (i.e., acoustic index window: 10 min; resolution: continuous except for gaps in recording). Recordings sometimes spanned different periods on different days (see above). Therefore, to obtain an average daily time series of 10 min resolution for each site, we assigned the acoustic index values to regular 10 min intervals (i.e., 00:00, 00:10, etc.) based on their start time (such that a value for e.g., 09:54 would be assigned to 09:50, and so forth). We then found the mean value for each 10 min period of the day present at each site, to obtain a representative diel time series of each index along the land-use gradient[7].

[5] A simplified version of the Loess detrending used for the satellite data.

[6] For the acoustic analyses, we calculate lag-1 autocorrelation directly, rather than fitting an order-1 autoregressive model and taking the coefficient, as for the satellite data. The latter has been applied both to vegetation satellite data [62] and other long time series of natural systems [107], motivating our use of it for assessing Bornean rainforest vegetation resilience. However, it assumes that the detrended windows of data are well-described by an order-1 autoregressive process (Section A.1.2). Given the more exploratory nature of our acoustics results, and potential imperfections from the detrending method, span and choice of window size, we opt to use functions that directly compute the correlation between a time series and a lag-1 shifted copy of itself instead.

[7] While the reassignment of index values to regular 10 minute intervals could lead to up to a 9 min

Since the old-growth forest recordings consistently only covered the daytime, we focussed on calculating EWSs over the period of 06:00 to 20:00 in the average diel time series for each site. We then performed approximate detrending by subtracting from each 10 min index value the mean for the 3 hr period of the day it fell within. Hence, a value for 09:50 would have the mean for 09:00-12:00 subtracted from it. Albeit less sophisticated than the regression-based Lowess/Loess methods used in other parts of the study, this simple approach has been explored in prior work [186] and is more appropriate given the short-term nature of the Borneo recordings and clear diel patterns in the indices. Lastly, we calculated variance and autocorrelation for each site on 3 hr windows that advance hourly. Existing studies have specifically shown the loss of characteristic avian choruses at dawn and dusk as a result of land-use change in tropical forests [26, 27]. We therefore expect examining the dynamics of these shorter fluctuations within the diurnal soundscape could yield insight into the stability of vocal activity patterns and resilience of sound-producing biodiversity to minor disturbances within the day. In the results, we present the mean values of the EWSs calculated over the average daily acoustic indices' time series for each site.

Robustness Tests

To test the robustness of our acoustics results for the Amazon, we explored varying the length of the sliding window the EWSs were calculated on to one week and four weeks. We also recalculated the results with a two-week sliding window using linear interpolation in the Python *pandas* package to fill gaps in the acoustic index data, rather than imputation with mean values. Lastly, we recalculated the results for one of the sites (Brazil Site 1) using a two-week sliding window on the acoustic indices' time series averaged down to 6 hr resolution (with the window thus advancing in 6-hr increments), to still capture diel trends but on a coarser dataset that might also capture longer-term fluctuations over multiple days.

For the acoustics results for Borneo, we recomputed the EWSs on 10 min periods using a 2 hr window advancing in 30 min increments and a 6 hr window advancing in 2 hr increments. We also recalculated the EWSs using the same 3 hr window advancing

discrepancy between a value's actual start time and assigned bin, an error of this magnitude is unlikely to occur frequently or make a notable impact on the averaged diel index time series (given the overall diel trend for the indices).

hourly as in the main text, but for the full averaged daily time series available for each site and for acoustic indices calculated on the first of every 10 mins of recording between 06:00 and 20:00 for each site (i.e., acoustic index window: 1 min; resolution: every 10 min).

5.3 Results

5.3.1 Satellite-Sensed Vegetation Resilience

Throughout the study period for our satellite data analyses, there was an upward trend in the mean AC(1) time series (Fig. 5.1), indicating a consistent reduction in rainforest resilience. A breakpoint occurred around March 2004 (95% Credible Interval (CrI): July

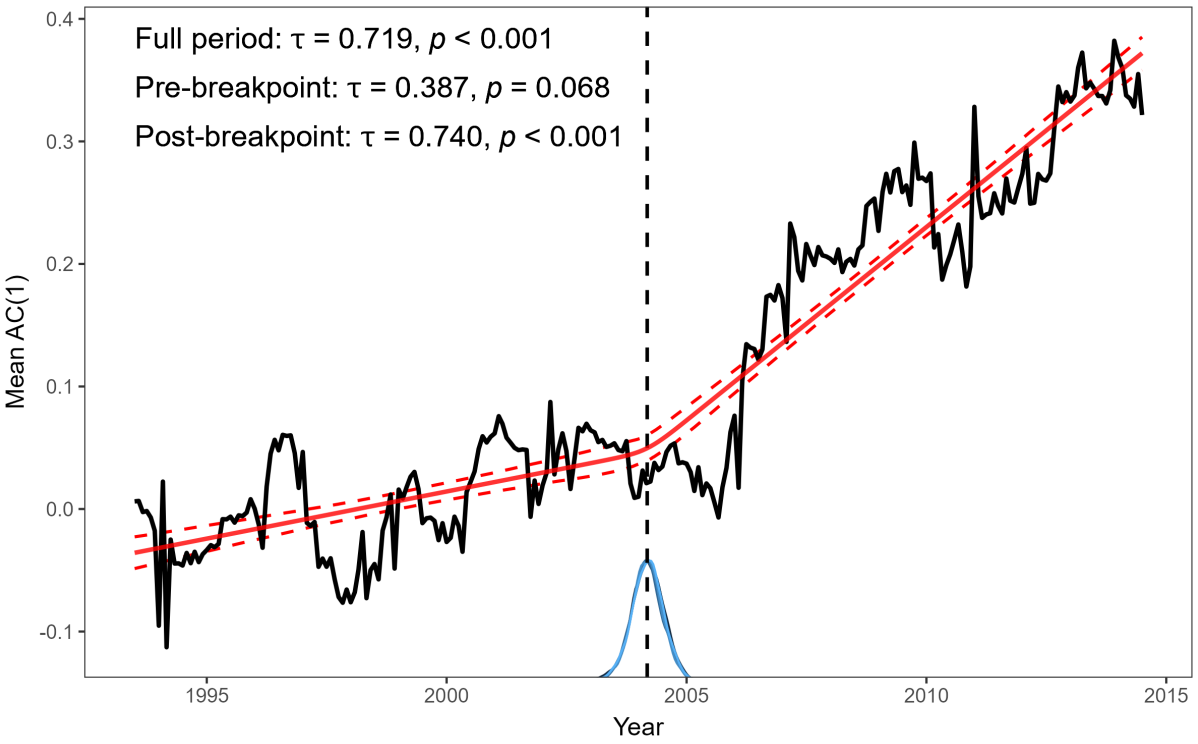


Figure 5.1: Overall satellite-derived vegetation resilience changes in Bornean rainforest, indicating consistent net resilience loss which accelerated from a breakpoint around March 2004. Mean Vegetation Optical Depth (VOD) lag-1 autocorrelation (AC(1)) time series of the study area from July 1993 to July 2014, calculated on 5-yr monthly sliding windows and plotted at the midpoint of each window. The red line represents the posterior mean from a Bayesian breakpoint regression; the red dashed lines indicate the 95% credible interval based on 50,000 posterior draws (10,000 iterations × 5 chains). The vertical dashed line indicates the posterior mean breakpoint, with the blue curve below showing its posterior distribution. Kendall's τ and associated p-values for the full record, pre-breakpoint, and post-breakpoint periods are reported in the top left.

2003–November 2004), after which the long-term upward trend became more pronounced (Fig. 5.1, Table D.1a). The timing of the breakpoint fell within a 5-yr VOD window that was potentially influenced by the additions of satellite sensors, AMSR-E (Advanced Microwave Scanning Radiometer - Earth Observing System) and WindSat, but it lies towards the end of this period and does not fully overlap with it (Fig. D.1). Further, we note that the mean VOD AC(1) also increased during a period without sensor changes, and sometimes decreased during periods with them, suggesting sensor changes are not the main influence on the trends in AC(1). Analysis of VOD variance also revealed an upward trend for the full and post-breakpoint time series (Fig. D.2). AC(1) and variance were positively correlated (Table D.2a), indicating that the observed changes in these indicators most likely reflect actual resilience loss (i.e., slowing down of fluctuations) [190].

Across Borneo, individual grid cells exhibited breakpoints around the same time as that observed for the entire study area (mean: February 2004[8], SD: 3.78 yr; Fig. 5.2a). Before the breakpoint, AC(1) increased in 52% of grid cells and across all grid cells the average Kendall's τ was 0.027 (SD: 0.47). After the breakpoint, AC(1) increased in 65% of grid cells and the average upward trend was more pronounced, with a mean Kendall's τ of 0.19 (SD: 0.47; Fig. 5.2b,c). There was a similar shift in AC(1) slope: 51% of grid cells exhibited a positive slope pre-breakpoint, rising to 67% post-breakpoint. Slope values were also strongly correlated with Kendall's τ (Fig. 5.2d,e, Table D.2a). Breakpoint timing was only weakly related to resilience loss before and after the breakpoint (Table D.2a). Notably, the consistency and magnitude of resilience loss in individual grid cells were inversely correlated between the pre- and post-breakpoint periods (Table D.2a), suggesting that areas that lost more resilience before the breakpoint went on to lose less after it, and those that initially lost less resilience went on to lose more post-breakpoint.

These results were generally robust to varying broadleaf forest cover thresholds used to determine the study area (Figs. D.3, D.4). For the overall trend, the AC(1) time series showed a similar upward trend across thresholds (60%: Kendall's $\tau = 0.71$; 90%: Kendall's $\tau = 0.709$; both $p < 0.001$; Figs. D.3a, D.4a). The lower 60% threshold produced a broader breakpoint posterior with an earlier main peak, suggesting overall resilience loss in forest including less intact areas may have accelerated from December 2000 (95% CrI: March

[8]Different to the Borneo-wide results reported above, as those overall results involve spatially averaging VOD before calculating AC(1) and variance, whereas here we are comparing the AC(1) of VOD time series for individual grid cells (Section 5.2.1).

1999–April 2004; Fig. D.3a). Increasing to 90% broadleaf cover yielded little change relative to the 80% baseline (July 2004, 95% CrI: December 2003–February 2005; Fig. D.4a). Spatially, the distribution of breakpoint timing and of pre- and post-breakpoint resilience changes (Kendall's τ and slope) remained similar (Figs. D.3b–f, D.4b–f). The results were also robust to different window lengths to estimate AC(1) (Fig. D.5).

For a more like-for-like comparison to the Amazon, we recalculated the VOD AC(1) result from Boulton et al. [62] using our processing pipeline (Section 5.2.1). The AC(1) time series for the Amazon still shows an increasing post-breakpoint trend (Fig. D.6, Table D.1a), with breakpoint timing remaining around early 2000 (Table D.1b). Across the Amazon, 82% of grid cells showed a change in their resilience trend later than the change for the overall area (Fig. D.6b), possibly indicating different mechanisms driving resilience loss compared to Borneo, where most grid cells passed their breakpoint around the same time as the aggregate area. The Amazon also showed apparent overall gains in resilience (declining AC(1)) before its breakpoint, and the consistency and magnitude of its resilience change were inversely correlated between the pre- and post-breakpoint periods (Table D.2b). In contrast, data for Borneo show resilience loss throughout the whole study period, though it accelerated post-breakpoint.

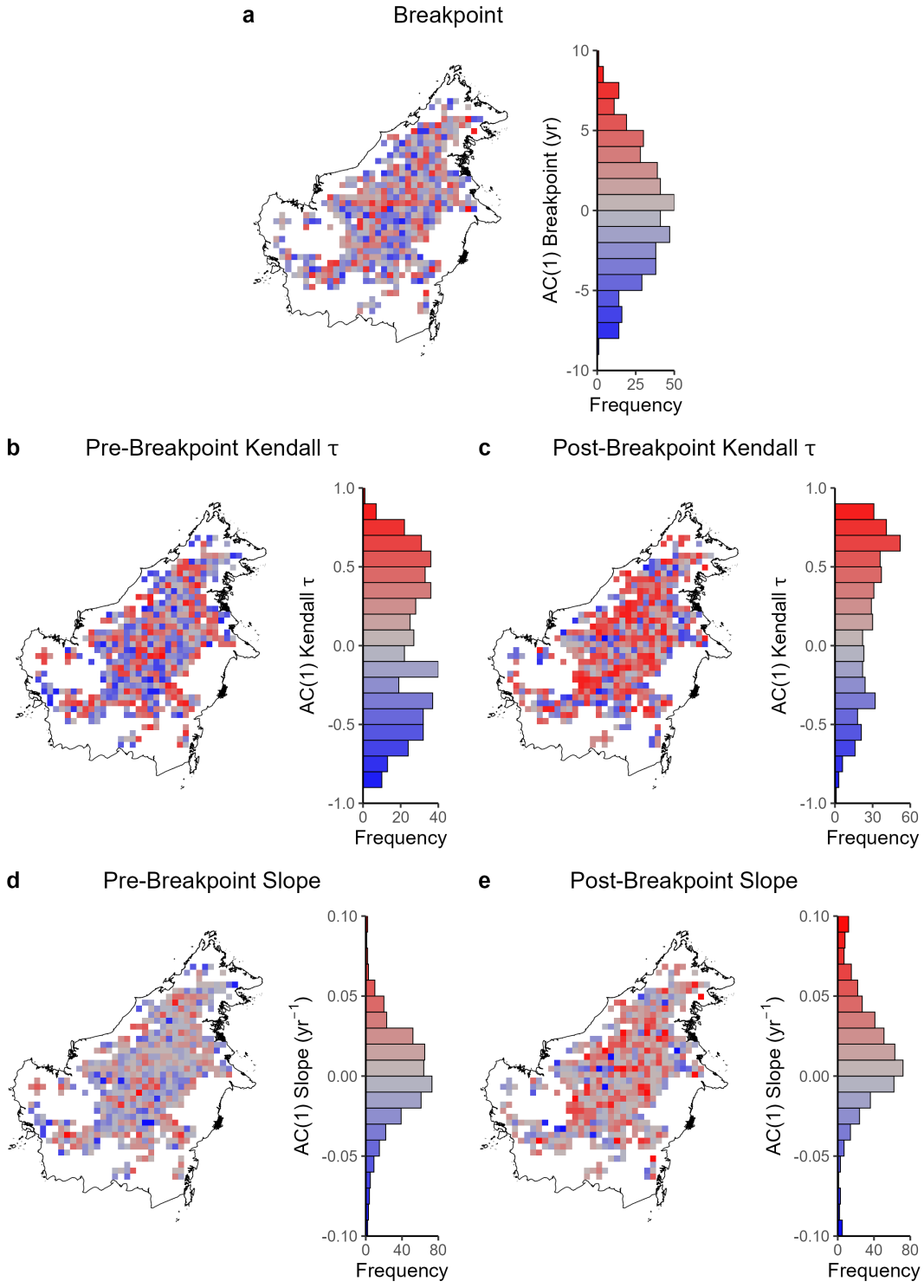


Figure 5.2: Spatial patterns of vegetation resilience changes in Bornean rainforest. Each panel shows a map of grid cells whose colours are related to the resilience metrics (breakpoint, Kendall's τ, slope) for the AC(1) time series of each individual cell. Vertical histograms serve as colour legends showing frequency distribution of grid cells' values. Breakpoint dates indicate when resilience loss accelerated, while Kendall's τ and slope respectively indicate the consistency and magnitude of resilience change (positive values indicate loss, see Section 5.2.1). (**a**) Map showing by how much the breakpoint date of AC(1) time series for individual grid cells deviates from that of the aggregate area (March 2004, Fig. 5.1). Positive values (coloured in red) show breakpoints after March

2004, negative values (coloured in blue) show earlier breakpoints. (**b**–**e**) Maps of grid cells' (**b**) pre-breakpoint Kendall's τ, (**c**) post-breakpoint Kendall's τ, (**d**) pre-breakpoint posterior mean slope (yr-1), (**e**) post-breakpoint posterior mean slope (yr^{-1}). Note, Bayesian quantities in (**a**), (**d**) and (**e**) are posterior means computed from 25,000 posterior draws (5,000 iterations × 5 chains). For visualisation, in (**d**) and (**e**), values outside the plotting range are constrained to the limits.

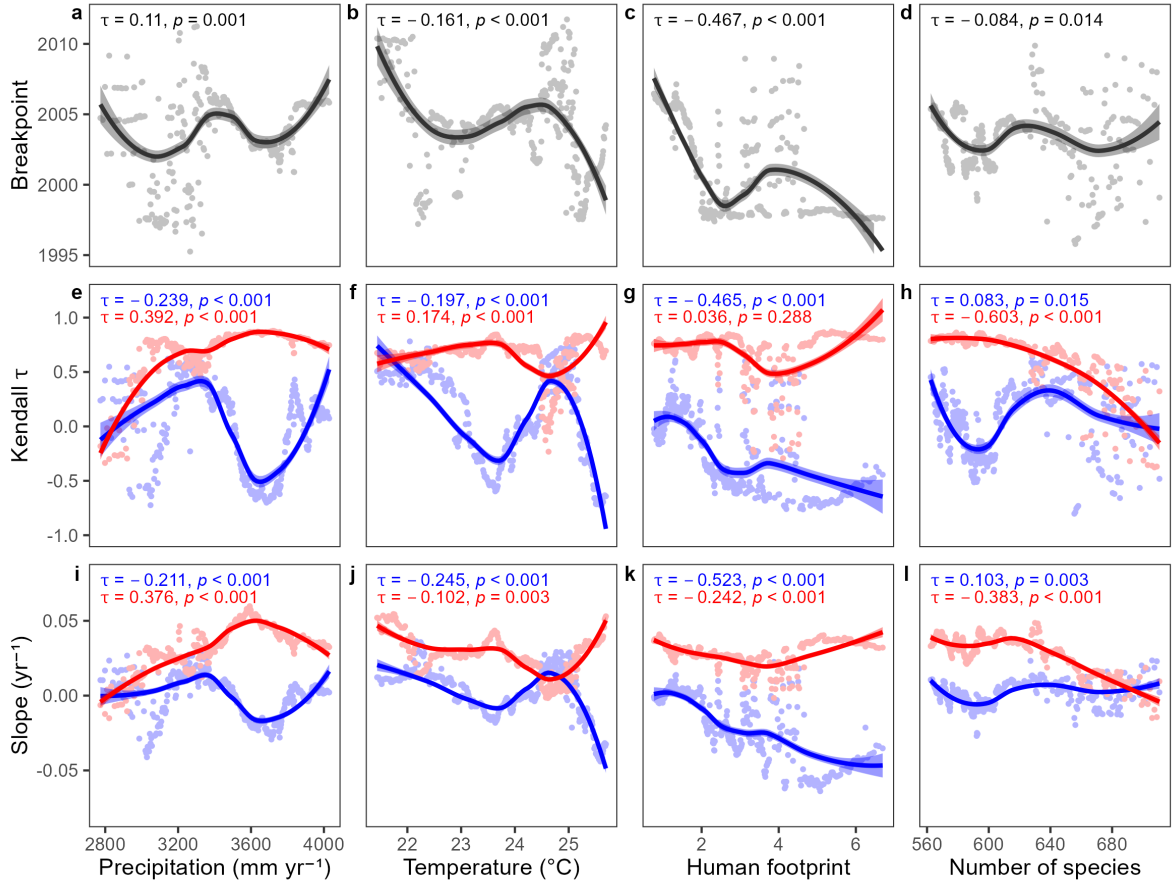


Figure 5.3: Relationship between vegetation resilience and potential influencing factors (sliding band size = 100 cells). Rows correspond to different resilience metrics (breakpoint – timing of accelerated resilience loss; Kendall's τ – consistency of resilience change; slope – magnitude of resilience change; Section 5.2.1) and columns represent different factors. (**a, e, i**) Correlation between precipitation and the following resilience metrics calculated on sliding bands of grid cells along a precipitation gradient (see Methods in Section 5.2.1): (**a**) posterior mean breakpoint from Bayesian breakpoint regressions; (**e**) Kendall's τ (blue: pre-breakpoint; red: post-breakpoint); and (**i**) posterior mean slope (blue: pre-breakpoint; red: post-breakpoint). The correlations between the resilience metrics and factors are calculated using Kendall's τ, which is reported with its p-values at the top of each panel. Remaining panels: the same as (**a, e, i**), but for temperature (**b, f, j**), human footprint (**c, g, k**) and animal species richness (**d, h, l**). Human footprint is a composite metric of human pressure on the environment [316] and species richness data are from the IUCN Red List [317]. Full sourcing and details for the factors are provided in Section 5.2.1. In each panel, the values are plotted at the midpoint of each band.

5.3.2 Correlates of Vegetation Resilience

The four factors we assessed for correlations to our resilience metrics (precipitation, temperature, human footprint and animal species richness) had weak pairwise correlations (Fig. D.7), reducing concerns about collinearity. Across all factors, breakpoint timing showed little relationship to each factor, except for human footprint (Fig. 5.3a–d), while the consistency and magnitude of resilience loss were generally higher after the breakpoint (Fig. 5.3e–l). Areas with lower precipitation had slightly earlier breakpoints to accelerated resilience loss (Fig. 5.3a). Drier regions were associated with more consistent and stronger resilience loss before the breakpoint (blue points; Fig. 5.3e,i), but wetter regions generally experienced more resilience loss after the breakpoint (red points; Fig. 5.3e,i). Temperature had an overall modest influence across metrics: breakpoints tended to occur earlier in hotter areas (Fig. 5.3b). Hotter areas experienced less resilience loss pre-breakpoint (blue points; Fig. 5.3f,j), but post-breakpoint associations between temperature and the consistency and magnitude of resilience change were mixed (red points; Fig. 5.3f,j). Interestingly, along both the precipitation and temperature gradients, pre- and post-breakpoint values of consistency and magnitude were negatively correlated (Table D.3a, Fig. 5.3e,f,i,j). In other words, the relationship between climate and resilience seemed to invert post-breakpoint. As human footprint increased up to a value of 2 units, areas exhibited earlier breakpoints, but the relationship between human footprint and breakpoint timing was relatively flat beyond this (Fig. 5.3c). Areas with more intensive human activity showed less consistent pre-breakpoint resilience loss (blue points; Fig. 5.3g) and lower magnitude resilience loss pre- and post-breakpoint (Fig. 5.3k).

For animal biodiversity, breakpoint timing (Fig. 5.3d) and pre-breakpoint resilience change had inconsistent relationships to overall species richness (blue points; Fig. 5.3h,l). However, post-breakpoint resilience loss was notably less consistent and pronounced in areas of high biodiversity after the breakpoint (red points, Fig. 5.3h,l). Individual taxa were intercorrelated to varying degrees (Fig. D.8), but the overall patterns in Fig. 5.3d,h,l were broadly consistent across groups (Fig. D.9). The overall trend of lower consistency and magnitude resilience loss in high richness areas was less pronounced for amphibians (red points; Fig. D.9e,i). Areas with high avian richness exhibited earlier breakpoints (Fig. D.9b). Unlike the pattern for overall biodiversity, resilience loss was stronger in areas of high avian richness and areas of high mammal richness (blue points; Fig. D.9f,j,g,k).

Reptiles largely mirrored the overall pattern (Fig. D.9d,h,l). The resilience change metrics along the overall species richness gradient had weak or non-significant correlations between their pre- and post-breakpoint values (Table D.3a). However, birds and mammals exhibited strong inverse correlations in their pre- and post- values of both Kendall's τ and slope, whereas amphibians showed positive pre– versus post-breakpoint correlations (Table D.3b). The above patterns across all factors and taxonomic groups were robust to the choice of grid cell band size (Figs. D.10, D.11, Table D.3).

5.3.3 Nascent Early Warning Signals in Acoustic Data

Both acoustic indices exhibited diurnal trends that changed between seasons in the Amazon (Fig. 5.4). On average, the acoustic complexity index had higher values in the daytime (from 06:00 to 18:00) with a peak around 07:00, slightly after the dawn chorus. The bioacoustic index followed an inverse trend, with higher values at night – when insect activity is highest – and a peak around 19:00, shortly after the dusk chorus. Across the day, the acoustic complexity index tended to have higher values in March (mid rainy season) and lower values in September (mid dry season). The opposite is again true of the bioacoustic index, which had higher values in September than March. Howells et al. [181] revealed broadly similar diurnal and seasonal patterns based on generalised additive models to predict acoustic index values from this dataset (albeit for the 3-10 kHz range and with a different data-cleaning procedure). Their study found the acoustic complexity index was highest in March and lowest in August, while the bioacoustic index was lowest in April and highest in October.

We find no clear trend in the EWSs calculated on the Amazon acoustic indices (Fig. 5.5). However, within each months-long section of data, we see some degree of correlation in AC(1) and variance for each index and between indices. This suggests that the EWSs are likely capturing changes to fluctuations in the acoustic indices and that these fluctuations change in a similar way for both the bioacoustic and the acoustic complexity index, either due to the indices describing similar soniferous biodiversity or components of biodiversity whose stabilities are evolving in a similar way. These results are generally robust to varying the sliding window length (although both the shorter one-week and longer four-week windows we tried produced slightly noisier patterns in the EWSs; Figs. D.12, D.13) and to using linear interpolation instead of imputation to fill gaps in the acoustic

indices (though patterns in the EWSs were somewhat altered; Fig. D.14). Resampling the acoustic indices for Brazil Site 1 to 6 hr resolution produced similarly inconclusive trends in the EWSs overall (Fig. D.15), but their patterns were notably different, with much lower values of variance and AC(1) (including some negative AC(1) values). These different patterns at 6 hr resolution were themselves fairly robust to different sliding window lengths (Fig. D.15).

In Borneo, we find that the average diurnal autocorrelation increases from old-growth to logged forest and is highest in oil palm for the acoustic complexity index (Table 5.1). The average variance for the acoustic complexity index similarly increases slightly from old-growth to logged forest, but is much lower for the oil palm plantation. This suggests slower fluctuations – and therefore potentially slower recovery from disturbances on the scale of minutes to hours – in soniferous biodiversity in logged versus old-growth forest. However, for the bioacoustic index, the average AC(1) and variance were both highest in the old-growth forest and lowest in logged forest, with only slightly higher values than logged forest in the oil palm. These patterns were robust to using both shorter and longer sliding windows and increments (Table D.4a, D.4b) and calculating the EWSs over the full average daily time series available at each site (as opposed to just the common 06:00 to 20:00 period; Table D.4c). Calculating the acoustic indices on the first of every 10 mins (to match the methodology for the Amazon acoustic indices) produced similar trends for the average variance, but the average autocorrelation progressively decreased from old-growth forest to oil palm plantation for both acoustic indices (Table D.4d).

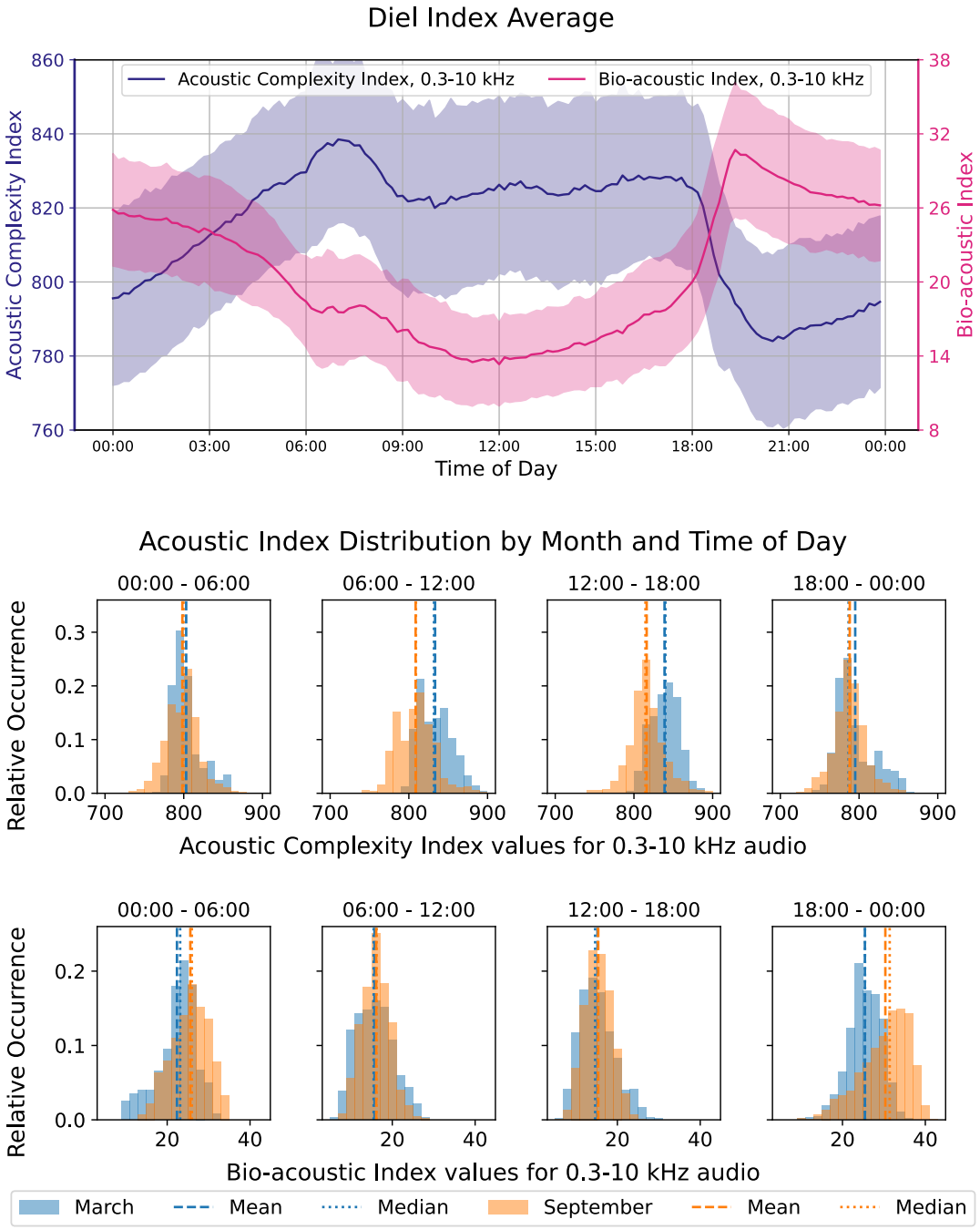


Figure 5.4: Diurnal patterns (top panel) and seasonal distributions (bottom two panels) of acoustic complexity and bioacoustic indices in recordings of the Amazon from Brazil Site 1. The top panel shows the mean values of the acoustic complexity and bioacoustic indices for every 10 min period of the day, based on the full recordings of Brazil Site 1. Shaded areas show standard deviation. The bottom two panels show histograms of the acoustic complexity (middle panel) and bioacoustic (bottom panel) indices for different periods of the day in March (blue bars) and September (yellow bars) 2018 and 2019, alongside the mean and median for each month and time of day (dashed and dotted lines, respectively; colour-matched to months). The histograms use relative occurrence (i.e., each bar is a proportion of total observations for its month and time of day) to better compare seasons, as recording gaps meant some months had more data points than others.

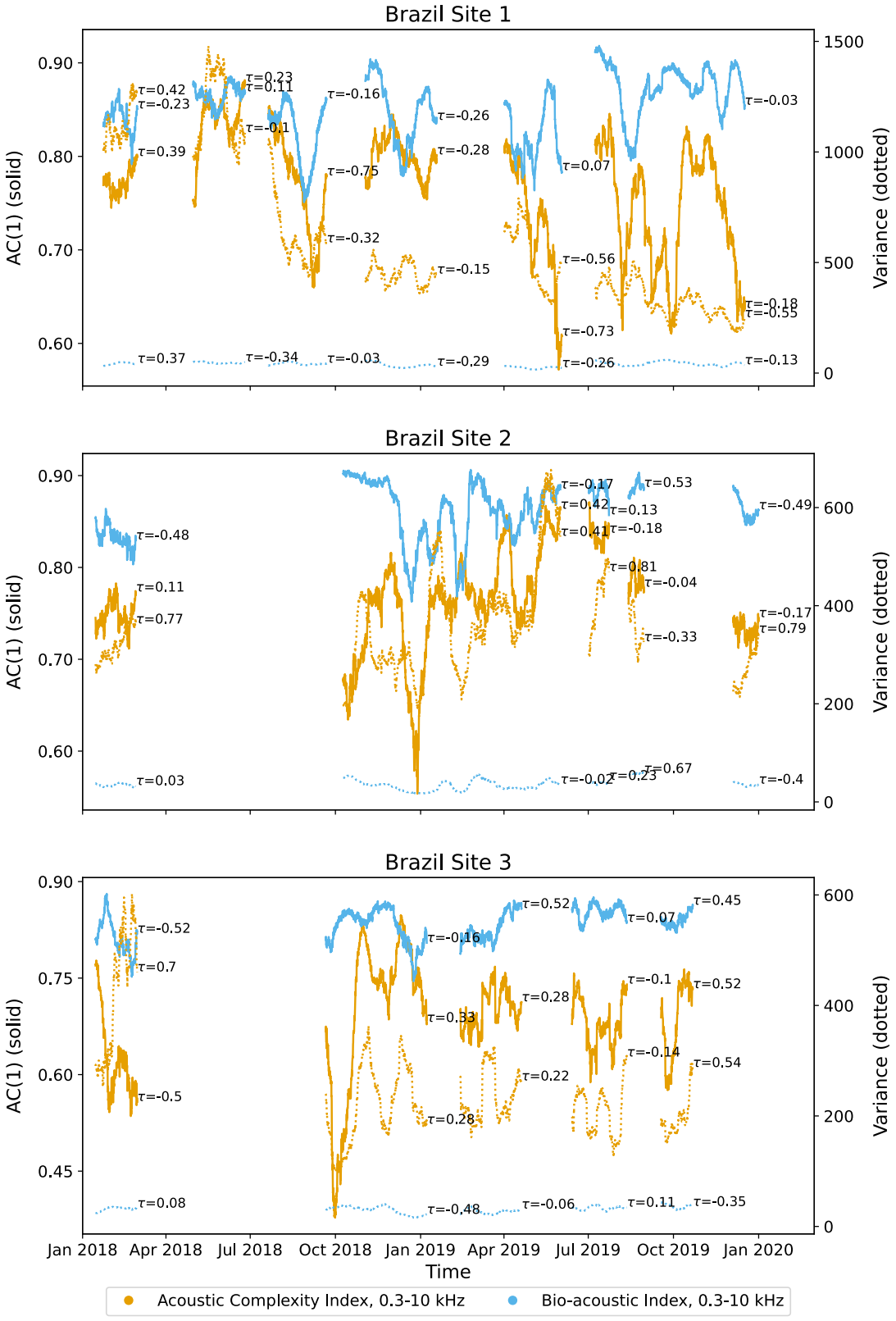


Figure 5.5: Early warning signals, lag-1 autocorrelation (AC(1); solid lines, left y-axis) and variance (dotted lines, right y-axis) calculated on detrended acoustic complexity and bioacoustic indices from recordings of the Brazilian Amazon using two-week windows advancing hourly (Section 5.2.2). Kendall's τ correlation coefficients indicate the consistency of trends for each months-long segment of data.

Table 5.1: Average values of early warning signals, lag-1 autocorrelation (AC(1)) and variance, calculated on acoustic indices from recordings gathered along a land-use gradient in Malaysian Borneo. The early warning signals are calculated on locally-detrended 3-hr-long windows of the acoustic complexity and bioacoustic indices that advance hourly in the daytime (Section 5.2.2). Abbreviations: Av - Average; Var - Variance.

		Old Growth	Logged	Oil Palm
Acoustic Complexity	**Av. AC(1)**	0.410	0.462	0.605
	Av. Var.	54.2	72.8	14.5
Bioacoustic	**Av. AC(1)**	0.713	0.521	0.587
	Av. Var.	4596	546	1272

5.4 Discussion

5.4.1 Satellite-Sensed Vegetation Resilience

We find that, similarly to the Amazon [62], Bornean rainforest vegetation is losing resilience. However, differences in the timing of increased resilience loss (breakpoints in AC(1); Figs. 5.1, 5.2, D.6) between these two regions suggest that this widespread phenomenon is likely driven by regional factors and/or local responses to global climatic events. While the Amazon switched from resilience gain to resilience loss around 2000, Bornean rainforest experienced continuous resilience loss throughout the study period (from the early 1990s), with an accelerated decline after 2004. The average rate of resilience loss in Borneo also appears to surpass that of the Amazon. This could mean tipping points are more imminent in Southeast Asia than in the Amazon, though we emphasise that the timing of tipping points depends on many factors and remains notoriously hard to predict.

Further differences between Borneo and the Amazon emerge when comparing landscape-scale and local timing of resilience change breakpoints. In Borneo, individual grid cell breakpoints generally occurred around the overall breakpoint. This implies that the Borneo-wide acceleration of resilience loss was a consequence of near-synchronous acceleration in resilience loss across subregions of forest. Conversely, in the Amazon, local-scale breakpoints typically occurred after the overall breakpoint. This may be due to regional-scale drivers – particularly severe climatic events such as widespread droughts – weakening resilience over large areas. This pattern may also be a result of gradual resilience loss in subregions of the Amazon exerting an outsized influence on other broad regions of the

forest. For instance, deforestation in upwind regions reduces evapotranspiration with severe consequences for downwind forests that lose an important share of their precipitation as a result [65]. More generally, at the local scale, breakpoint timing was only weakly related to the degree of resilience loss before the breakpoint, suggesting that increases in resilience loss are not necessarily triggered by the degree of prior loss.

Although Borneo lost resilience overall and within most of its subregions, a number of grid cells did not lose resilience or even appear to have gained it. Further, in both Borneo and the Amazon, places that lost more resilience before the breakpoint tended to lose less – or even gain resilience – afterward, and vice versa (Table D.2a). While further work is needed to understand this pattern, it reveals that big-picture, landscape-wide resilience loss arises from a mosaic of local patterns and factors that can be asynchronous and interact in various elaborate ways, including this apparent reallocation of resilience between subregions. More generally, this finding underscores the complexity and scale-dependent nature of ecosystem resilience [328–330].

5.4.2 Correlates of Vegetation Resilience

Across Borneo, the onset of accelerated resilience loss tended to occur earlier in drier and hotter regions. Such areas have been shown to exhibit lower functional diversity, declining functional redundancy and heightened vulnerability, potentially predisposing them to earlier acceleration of resilience loss [331, 332]. Grid cells with lower precipitation lost more resilience before the breakpoint, broadly comparable to prior findings [62, 63, 113]. However, after the breakpoint, wetter regions exhibited more pronounced resilience loss, perhaps because high precipitation can become a stressor once forest condition is compromised by other factors (e.g., via waterlogging and root hypoxia and/or heightened pathogen and pest pressures; [333–335]). By contrast, no clear patterns in pre- and post-breakpoint resilience loss emerged along the temperature gradient, perhaps because temperatures largely fell within a suitable range and were not a key stressor in Borneo. The reversal of patterns between climate and resilience after the breakpoint may account for the apparent spatial reallocation of resilience post-breakpoint described earlier. However, further work is needed to understand whether and why this might be the case, as well as to explore whether these findings are an artefact of the breakpoint method and decision to treat factors as constant. Indeed, a simple explanation could be that precipitation and

temperature were different for the pre-breakpoint period.

Regarding human activity, areas with higher human footprint tended to reach the breakpoint earlier, consistent with anthropogenic disruption of key ecological processes and networks, such as seed dispersal and pollination, which can increase vulnerability to disturbances [304, 336, 337]. Human-induced fires and deforestation can also produce a drier microclimate that makes it easier for subsequent fires to start and spread [63, 65]. In contrast to findings from the Amazon [62], we observed that areas with greater human footprint were associated with less pronounced loss of resilience overall. This could be indicative of positive human influences, such as reforestation and forest restoration efforts, though this interpretation requires further validation.

For biodiversity, we found that areas with higher species richness lost less resilience after the breakpoint. This pattern appeared for overall animal biodiversity and, broadly, within individual animal taxa. As outlined in the chapter introduction, animal biodiversity can contribute to tropical forest resilience through mechanisms like seed dispersal [304]. Further, previous research suggests high species richness enhances functional diversity, allowing disturbances to be partitioned or mitigated through asynchronous and variable species responses, while functional redundancy ensures that ecosystem processes remain intact when certain species are threatened, ultimately bolstering overall resilience [223, 338, 339]. On the other hand, although short-term resilience loss might not immediately trigger a regime shift, it may manifest in more subtle ecosystem impacts (e.g., leaf wilting, fruit scarcity), prompting animals to migrate to more resilient areas in search of resources. Indeed, the acoustics fieldwork in Borneo (September 2024) overlapped with the fruiting season, during which unusually high numbers of orangutans (up to seven) were noted in some individual trees near the base camp rest houses in old-growth forest. Local research assistants noted this might be due to hotter and drier conditions under climate change reducing the availability of fruit and/or disrupting its distribution. However, the IUCN biodiversity data used in our satellite analysis here is aggregated over decades and lacks population-level resolution, hence we cannot conclusively disentangle the mechanisms behind these Borneo-wide patterns. Our acoustics results, though, demonstrate a potential route forward for using high-throughput autonomous monitoring to gain the more fine-grained insights needed to unravel how biodiversity and forest-wide resilience interact.

Like climate, biodiversity appears to modulate the reallocation of resilience loss post-breakpoint, but the effect is taxon-specific. Birds and mammals both showed the clearest reversal in their pre-/post-breakpoint trends between richness and resilience, such that post-breakpoint, lower richness areas had more pronounced resilience loss and higher richness areas experienced less resilience loss or even resilience gain. This may be reflective of biodiversity's broader ability to enhance response diversity and/or buffer shocks through asynchronous species dynamics [223, 338]. By contrast, amphibians displayed more similar pre/post relationships to resilience, while reptiles showed less consistent pre/post patterns. These contrasts may arise from differences in mobility, habitat breadth, and ecological roles. In particular, birds and mammals typically traverse larger spatial mosaics and often occupy higher trophic positions that can exert top-down influences on community dynamics [223, 302, 338]. They may also reflect variation in regional species-pool sizes, which affects the detectability of richness effects in our spatially aggregated data. Disentangling these pathways will require community-level studies that integrate resilience dynamics with interaction-network structure and functional roles.

5.4.3 Nascent Early Warning Signals in Acoustic Data

Our acoustics results were in large part inconclusive. The lack of clear EWS trends in the Amazon and discrepancies in findings between the Amazon and Borneo could stem from the inherently local nature of soundscapes. However, these inconclusive patterns more likely reflect the different timescales of fluctuations we studied in each region (i.e., *of* versus *within* diel fluctuations) and the need for data of even greater spatiotemporal scale to robustly identify EWSs in the acoustic domain, should they exist. Although we analysed comparatively long terrestrial passive acoustic monitoring time series from the Amazon, they were limited both in length (less than one tenth of our satellite time series) relative to the decadal timescale of forcing from climate change, and in resolution, given the long gaps in data collection. To circumnavigate the challenge of temporally-limited recording in Borneo, we effectively performed a space-for-time substitution, allowing us to look at shorter-scale soundscape fluctuations in progressively disturbed environments. However, longer term data would also have been useful here to look at fluctuations on the scale of months for more direct comparison to shifts in satellite-sensed vegetation resilience, in addition to the short intra-diurnal fluctuations we examined. Our recordings

also correspond to a small spatial scale. Site replicates and a more extensive and granular disturbance gradient would be needed to more rigorously explore spatial patterns in acoustic signs of resilience loss, and compare them to satellite data and the climate and human activity factors we explored here. Longer-term ecoacoustic datasets of ten years or more do exist in the marine realm (see references within [117]), and are beginning to emerge for terrestrial ecoacoustics (e.g., [45]), as are terrestrial deployments over larger spatial scales [124–127]. We also hope passive acoustic monitoring with MAARU (and its future iterations) on larger spatiotemporal scales will be possible with some of the deployment recommendations we shared in the previous chapter.

Our analysis pipeline is generally in line with the analytical framework we proposed for acoustically monitoring tipping points in Chapter 2, with the exception of step 2, to gather data of sufficient scale and resolution. Step 1, to identify tipping potential, is met since tropical rainforests are known to possess alternative grassy and treeless states (Sections 1.1.3 and 5.1). For step 3, to extract relevant ecological metrics from data, we examined two well-established acoustic indices that have been studied for their relationship to biodiversity in our Amazon dataset and other similar data previously [144, 181]. We filtered the frequency range of the audio prior to calculating the indices to limit self-noise and other less relevant sounds below 300 Hz and provide an acceptable signal-to-noise ratio for a wide range of taxa in Borneo and the Amazon. We found expected seasonal and diurnal patterns in the acoustic indices, which informed our approaches to resampling and de-trending the data, per step 4 to perform pre-processing. Future work could examine different resampling and de-trending timescales and/or more restrictive frequency filters to study more specific subsystems or particular species. Finally, for step 5, to search for EWSs, we searched for variance and autocorrelation, in line with our satellite-based analyses and the majority of research on EWSs of tipping points [60].

Our robustness tests showed that, for the most part, our results do not appear to be an artefact of specific methodological choices and parameters. However, the results of resampling the Brazil acoustic indices to 6 hr resolution (Fig. D.15) suggest that the higher autocorrelation and variance for the results at hourly resolution arise from diel fluctuations in the acoustic indices appearing slower (i.e., spanning more data points) at this finer resolution. Whether monitoring fluctuations at either of these temporal resolutions could indicate impending biodiversity collapse, and which one would be most suitable

for doing so, merits further investigation. Furthermore, the diel acoustic index fluctuations we probed for slowing down could be interpreted as stable oscillatory avian singing dynamics, rather than noise or perturbation-induced variation in a proxy of soniferous biodiversity abundance. This may underpin why the examination of fluctuations *within* diurnal soundscape patterns in Borneo revealed a clearer trend of slowing down between old-growth and logged forest. However, the patchiness of our Amazon recordings limited the extent to which we could study fluctuations in acoustic indices on finer or coarser scales. Moreover, our approach is appropriate for tackling the open question of whether such diel fluctuations are in fact slowing down as pressures on the Amazon intensify. We also did not investigate the impact of the Lowess detrending span here. Therefore, further robustness testing – especially of pre-processing parameters – and larger-scale data are the missing pieces to exploring what scope the above framework has for effectively complementing satellite data in forewarning of ecosystem tipping points.

Nonetheless, our results show some promise for acoustically monitoring tipping points. First, the similar patterns in AC(1) and variance within and between acoustic indices in Brazil provides initial validation that our processing pipeline is indeed capturing the dynamics of fluctuations in acoustic descriptors of biodiversity. Monitoring these fluctuations on a longer timescale could therefore reveal critical slowing down in response to climate and anthropogenic pressures, should it occur. Further, the increase in both AC(1) and variance of the acoustic complexity index in logged forest in Borneo reveals potential critical slowing down of soniferous biodiversity's recovery from short-term perturbations within the day. However, this trend was not observed for the bioacoustic index. This might be because the bioacoustic index is predominantly capturing insect activity, given its higher nighttime values and ability to best predict insect acoustic activity in Howells et al.'s study [181]. With its higher daytime values, the acoustic complexity index could, on the other hand, be better capturing avifauna and other animals which might be more directly related to the resilience and quantity of vegetation. The acoustic complexity index could, for example, capture the vocal activity of frugivores, which can support rainforest resilience through fruit consumption and seed dispersal [304] but are being adversely affected by disruptions to fruiting from climate change (e.g., observational evidence of orangutans finding fewer fruit-rich trees in Borneo, as mentioned in Section 5.4.2). However, while logged forest is home to less biodiversity [287] and a loss

of temporal soundscape structure compared to old-growth forest [26, 27], the question of whether slowing diurnal fluctuations in vocal activity are a precursor of a tipping point in soniferous biodiversity or the rainforest at large remains open to debate. Addressing this debate again requires further data; more complex ecosystem models that explore biodiversity dynamics in relation to land-use and vegetation resilience may also be useful.

5.5 Outlook and Conclusions

By introducing two additional metrics of resilience change, breakpoint and slope, we captured more detail on the correlates and potential drivers of resilience loss in Bornean rainforest vegetation. Further investigation is needed to determine whether these relationships are causal. Factors beyond those considered here, as well as indirect or multivariate effects, likely also shape the dynamics and resilience of complex ecosystems [60, 94, 221]. However, those we examined are considered to be the predominant global and local factors impacting rainforests worldwide [62–64, 113, 223, 304, 338, 339].

As highlighted in Chapter 2, EWSs derived from critical slowing down can indicate resilience loss and increasing proximity to a possible tipping point, but cannot accurately predict the type of impending critical transition or when it might occur [72]. Further, critical transitions may occur without critical slowing down [191, 340, 341] and critical slowing down may also produce false alarms [72, 189]. However, as in other studies, we substantially reduced the likelihood of these errors by examining AC(1) and variance, and finding similar patterns of resilience loss in both. Moreover, these EWSs remain useful in suggesting an increased risk of forest instability and vulnerability to hazards, which could offer actionable guidance to Bornean conservation initiatives, such as the Heart of Borneo [342].

For a better understanding of whether, how and when tropical forest ecosystems could tip, future work could explore recent deep learning time series classifiers that can indicate the type of bifurcation – if any – that lies ahead [187], while larger-scale passive acoustic monitoring deployments could provide observations towards unifying animal biodiversity and vegetation resilience and their mechanisms of interaction. Further integrating satellite remote sensing and fine-grained, high-throughput ecosystem monitoring technologies like acoustics could also provide multi-scale insights into ecological resilience. Using such

multimodal monitoring data to build and refine holistic ecosystem models (e.g., [217]) may also allow for ecological resilience and tipping points to be forecast under different scenarios of future climate, land-use and biodiversity change.

In summary, we found that long-term resilience loss in rainforests is a pan-tropical issue, but it appears to be shaped by spatial variations in biotic and abiotic factors. These insights may inform scientific guidance for the conservation and management of rainforests, especially in terms of addressing climate change, managing human activities, and protecting biodiversity. Moreover, our findings indicate that regional variation in key global drivers of rainforest resilience loss (e.g., global warming) and local factors interplay to cause differences in the underlying mechanics, breakpoint, consistency and magnitude of Bornean versus Amazonian rainforest resilience loss. Our exploratory acoustics results suggest a possible pathway forward to gaining deeper understanding of these local processes. We found that generic statistical EWSs may also capture changes to fluctuations in acoustic biodiversity metrics and detected slower diurnal soundscape dynamics in logged versus old-growth Bornean forest, indicative of potential resilience declines in soniferous biodiversity – though further work is needed to validate if this is the case. These methodological advancements highlight the potential for multimodal resilience monitoring that incorporates satellite and acoustic data to better understand and anticipate ecosystem tipping points. Collectively, our expanded evidence of resilience loss towards a possible tipping point of mass forest dieback in Borneo emphasises the urgent need for cross-scale conservation of rainforests, including their animal biodiversity, and international climate change mitigation.

Chapter 6

Conclusion: Raising alarms and sounding change

6.1 Synthesis

In this thesis we have described the development and investigation of new approaches across the ecoacoustic monitoring pipeline to assess the spatial and resilience-related features of natural soundscapes.

We began by reviewing tipping point and resilience theory and detection alongside cutting-edge developments in autonomous ecoacoustic monitoring in **Chapter 2**. From this, we presented a promising methodological framework for acoustically monitoring ecosystem tipping points across scales. Our literature review also highlighted the importance of carefully calibrating and understanding the performance of ecoacoustic monitoring tools, from recording hardware to analytical methods. In **Chapter 3**, we therefore shifted focus to the testing of (spatial) ecoacoustic technologies. We built and pioneered the use of an ambisonic Virtual Sound Environment (VSE) to simulate natural soundscapes for evaluating multichannel recording hardware (MAARU) and software for bird call classification (BirdNET) and localisation (HARKBird).

Building on this, in **Chapter 4**, we moved from testing MAARU in a VSE to real field environments in the UK, Brazil, USA and Malaysia, in order to investigate its robustness and capacity to target or distinguish sounds from different directions by beamforming (directional filtering). As part of this effort, we redesigned MAARU's weatherproofing and field deployment setup to be easier to build, provide better access to internal elec-

tronics and powering components, and be more durable to challenging environmental conditions (e.g., high temperature and humidity, animal interference) ahead of longer-term deployments in the tropical forests of Malaysian Borneo. While several devices still did not survive past two weeks owing to new failure modes we discovered – and in some cases addressed – across various deployments, we showed that the confidence of BirdNET predictions made on omnidirectional audio can be significantly improved by simple directionally static beamforming. Furthermore, we demonstrated this scalable beamforming strategy can also return more BirdNET predictions overall – perhaps by capturing more distant vocalisations – and can reveal patterns in the spatial distribution of biodiversity, including those that may arise from habitat or landscape structure.

In **Chapter 5**, we explored using satellite and acoustic data to detect EWSs of tipping points at different ecosystem levels and temporal scales in tropical rainforests. Results from decades of satellite monitoring data revealed worrying declines in Bornean rainforest resilience that accelerated from the early 2000s, as was also the case for the Amazon. In studying how spatial patterns of this decline relate to various factors that may shape resilience, we uncovered that subregions with greater species richness lost less resilience, and birds and mammals in particular might even play a role in reallocating resilience loss between areas – though further work is needed to validate this claim. However, EWSs calculated on acoustic biodiversity indices from recordings over a two-year period in the Amazon and week-long recordings along a land-use gradient in Malaysian Borneo were largely inconclusive. These findings reveal that even larger-scale data and further analyses are needed to determine whether EWSs of resilience loss towards tipping points can manifest in the acoustics domain and be detected in soundscape metrics. We still propose searching for acoustic EWSs could be fruitful, though, as our results also show preliminary evidence that generic EWSs (autocorrelation and variance) appear to capture fluctuations in aspects of sound-producing biodiversity, with slower intra-diurnal dynamics in human-modified compared to intact rainforest in Borneo.

6.2 Evaluation of Research Objectives

In relation to the Research Objectives we set at the start of this thesis, our work from **Chapter 3** largely satisfied Objective 1, to develop a representative, controlled and re-

peatable means of assessing spatial ecoacoustic technologies. However, we also found limitations to using an ambisonic VSE to this end, mainly because of its bias to human perception. We thus shared guidelines on the appropriate construction and use of ambisonic VSEs for testing ecoacoustic technologies, and suggested an alternative approach of sound field reconstruction (vector-based amplitude panning) if more accurate spatial rendering is required. In **Chapter 4**, we satisfied Objective 2, revealing that beamforming can enhance biodiversity monitoring. However, additional analysis could further strengthen and/or explain our results. For instance, further analysis of spectrograms could confirm whether BirdNET returns more predictions on beamformed audio as a result of fainter bird vocalisations being amplified across the diverse ecosystems we studied. To better understand to what extent beamforming can spatially segment soundscapes, another interesting next step would be to compare acoustic indices and deep learning embeddings on omnidirectional audio and recordings beamformed to different directions. While we were unable to fully realise the framework for acoustically monitoring tipping points we devised in **Chapter 2**, our work in synthesising broad literature and theory on this topic and in examining satellite and acoustic data for EWSs based on critical slowing down in **Chapter 5** represent a significant contribution towards Objective 3. Further, our work could be extended with more robustness tests, including variation of the detrending methods used, and by searching for EWSs in additional ecoacoustic metrics, such as the predictions and embeddings of bioacoustic classifiers like BirdNET.

In addition to the short-term next steps for this work outlined above, this thesis also lays the groundwork for future research directions, which we summarise below.

6.3 Future Directions

6.3.1 Where Next for MAARU?

Over the course of this PhD, we have had interest from researchers wishing to use MAARU in the UK, the Netherlands, India, Costa Rica, New Zealand and China. We have also supported work to install MAARUs in the ecological monitoring system for the new gardens of the Natural History Museum, London. Clearly, the appetite for autonomous spatial ecoacoustic monitoring is substantial. However, MAARU faces a major setback in that *Seeed Studio* have ceased production of the ReSpeaker 6-microphone array used

in the device, and few alternative options are available. In response, we have initiated a collaboration with researchers and engineers at the Cornell Lab of Ornithology's K. Lisa Yang Center for Conservation Bioacoustics to design a new custom MEMS microphone array that could power the next generation of MAARU. This work is ongoing, and could pave the way for enabling low-cost multichannel ecoacoustic monitoring at scale.

In addition, ongoing work led by fellow PhD student, Rifqi Ikhwanuddin, is testing more sophisticated beamforming methods – including a technique based on impulse responses measured at multiple locations – in an updated version of our VSE. This work involves using the VSE to precisely vary signal-to-noise ratios to explore how beamforming with MAARU can enhance progressively degraded signals, highlighting further ways for VSEs to benchmark ecoacoustic tools.

6.3.2 Realising Multimodal Tipping Point Monitoring with (Spatial) Ecoacoustics

Despite our acoustics results in Chapter 5 proving largely inconclusive, our review and proposed analytical framework in Chapter 2 point to major opportunities for future research on using acoustics to anticipate tipping points and integrating it with other sensing modalities. As in many fields, deep learning approaches across the whole pipeline seem like a particularly promising avenue, from the extraction of ecological information from soundscapes to the prediction of bifurcations from time series. In addition, hidden Markov models could provide a way to predict regime shifts using monitoring data that are spatially extensive but temporally sparse. Such approaches that rely on large spatial rather than temporal coverage are especially relevant to recent passive acoustic monitoring deployments across vast geographic areas that have yet to gather long-term data. Acoustics may also be particularly well-suited to detect vocal signatures of trait-based changes in populations approaching collapse. More focussed acoustic studies that examine particular species known to exhibit such trait changes could thus be particularly useful for expanding the toolkit of possible acoustic EWSs that do not rely on critical slowing down.

This thesis has also provided initial evidence that spatial ecoacoustic technologies could complement and enhance the above methods in various ways. For instance, beamforming with multichannel recorders could improve the performance of species call classifiers by reducing background noise and masking, leading to better assessment of species richness

and population dynamics. By focussing on target signals, it might also allow for more accurate quantification of trait change signatures in vocalisations. In addition, beamforming with multichannel recorders may reduce costs in large-scale deployments by allowing individual recording units to distinguish neighbouring sites. This could reduce the number of units that need to be deployed to cover extensive areas or ecological gradients, offering a route to large-scale monitoring with more restricted research budgets. Estimating the direction of arrival of vocalisations with multichannel recorders can also reveal the distribution of soniferous biodiversity around recording units, and time-synchronised networks of such units could be used to precisely track the position and movement of vocal animals. While these spatial ecoacoustic technologies merit further research and development (e.g., superior microphone arrays and beamforming techniques and validation of their impact on ecoacoustic analyses), they could bridge the spatial and resilience-related features of natural soundscapes by providing novel or improved insights into the spatial and temporal dynamics of biodiversity that support tipping point assessment.

For all the promise ecoacoustics holds for anticipating tipping points, its strengths and impact will be greatest when used alongside other monitoring approaches. In particular, acoustic monitoring might at present be best suited to complementing established, large-scale satellite monitoring of ecosystem tipping points through targeted deployments in hotspots of declining resilience. These localised acoustic monitoring efforts could then provide additional detail on the nature of resilience loss. To realise multimodal tipping point monitoring, future research could, as a first step, compare EWSs of tipping points in a given system that are derived from different sensing modalities. For instance, one could compare EWSs in vertebrate population dynamics estimated from cameras and acoustic sensors. Alternatively, given soundscapes carry information about vegetation (through e.g., acoustic reflections and the presence of vocal animals attracted to certain plants), one could explore whether coarser-resolution (e.g., monthly) soundscape metrics and satellite vegetation indices exhibit EWSs that correlate over long time periods. In turn, it would be interesting for future research to investigate whether EWSs derived from sensors that monitor ecosystems at different levels also correlate, in order to understand how resilience loss manifests across different variables within complex natural systems. For example, one could explore whether animal biodiversity in tropical rainforests shows declining resilience as vegetation in those regions exhibits critical slowing down. In ad-

dition, ecoacoustic insights into spatiotemporal biodiversity dynamics could also unravel potential mechanisms of impending tipping points detected with other forms of sensing. Beyond these approaches, data from different sensing modalities including acoustics could be fed into more advanced multimodal ecosystem models or multivariate EWSs in future.

Integrating ecoacoustics into truly multimodal, real-time monitoring pipelines could prove invaluable for understanding the mechanisms and present risk of resilience loss across scales. To maximise the ecological insight and conservation benefits of ecoacoustics will also require more broadly situating findings from acoustic monitoring, complementary sensing technologies and the Western scientific method alongside other established systems of ecological knowledge, management and understanding. In particular, there has in recent years been a growing appreciation within academia of local and Indigenous knowledge, which has allowed for the sustainable co-existence of humans within ecosystems for thousands of years [343, 344]. Combining ecoacoustics research with local and Indigenous knowledge systems to avert tipping points and deliver effective conservation can take many forms and will naturally be geographically-, culturally- and otherwise context-specific [343]. One approach is for ecoacoustic experts to empower local communities and scientists with ecoacoustic tools and training, allowing for locally-led monitoring and conservation initiatives. This approach has been successfully implemented in the BEAT (Bioacoustics Equipment And Training) programme in Malaysia and Indonesia [345] and the Elephant Listening Project in Central Africa [346], both led by the K. Lisa Yang Center for Conservation Bioacoustics. Ultimately, leveraging ecoacoustics with other sensing modalities and alongside or within local and Indigenous knowledge systems could allow for the rapid identification and targeted conservation of at-risk ecosystems before they reach critical transitions.

6.3.3 Positive Tipping Points

Tackling the climate and ecological crises can feel like a Sisyphean task. But perhaps the goal is not to push a boulder to the top of the mountain, but to get it to the other side – rather like pushing a ball in a potential landscape from one basin of attraction to another. This analogy gets to the heart of *positive* tipping points: social and technical movements that, after gaining traction, eventually become self-sustaining (analogous to media or ideas 'going viral') and can drive runaway societal change [5]. Examples of

such movements include the development and adoption of new sustainable technologies like renewable power and electric vehicles, and social or behavioural movements, such as the historic Suffragette and Civil Rights movements. More recent socio-behavioural examples include the take-off of Extinction Rebellion in the late 2010s and the growth of plant-based diets.

It is now argued that such transitions to more sustainable ways of living are among our best bets at curbing the fast-approaching negative tipping elements in the Earth system. Further, the same generic EWSs of dangerous climate and ecosystem tipping points are now also being identified as 'early opportunity signals' in data of beneficial social-technical movements. For example, Boulton et al. [347] recently found rising early opportunity signals in a time series of searches for electric vehicles on a popular UK car sales website. This is just one example of many such signals that are emerging [1].

The message of positive tipping points should be one of hope: whilst efforts to make real progress on climate and environmental change have been gruelling, and may currently face setbacks, we might be close to reaching rapid, self-amplifying change. With small further pushes, gravity could soon bring the boulder down the other side of the mountain. However, we argue that to build such movements in a way that is resilient to setbacks at a societal level also requires resilience at the individual level. In this context, though, the current predicament of the Earth system poses further challenges through its detrimental impact on individual mental and emotional wellbeing [348] (sometimes termed 'eco-anxiety').

We posit natural soundscapes might play a role in building both individual and societal resilience. We know that hearing natural soundscapes positively impacts human wellbeing (building individual resilience) and fosters a deeper sense of connection to nature (building pro-environmental social movements) – particularly in urban settings [21, 349]. Soundscapes can also be at the heart of engaging the public with climate and ecological research, through citizen science initiatives that use user-submitted field recordings to track ecological patterns [143, 350]. Therefore, we conclude that as well as their potential for helping understand the fate of biodiversity and ecosystems, the emotional and social potency of natural soundscapes may be amongst their most valuable resilience-related features for safeguarding the future of life on Earth.

Bibliography

[1] T. M. Lenton, M. Milkoreit, S. Willcock, J. F. Abrams, D. I. Armstrong McKay, J. E. Buxton, J. F. Donges, S. Loriani, N. Wunderling, F. Alkemade, M. Barrett, S. Constantino, T. Powell, S. R. Smith, C. A. Boulton, P. Pinho, H. Dijkstra, P. Pearce-Kelly, R. M. Roman-Cuesta, and D. Dennis, Eds., *The Global Tipping Points Report 2025*. Exeter, UK: University of Exeter, Oct. 2025.

[2] D. I. Armstrong McKay, A. Staal, J. F. Abrams, R. Winkelmann, B. Sakschewski, S. Loriani, I. Fetzer, S. E. Cornell, J. Rockström, and T. M. Lenton, "Exceeding 1.5°C global warming could trigger multiple climate tipping points," *Science*, vol. 377, no. 6611, p. eabn7950, Sep. 2022.

[3] A. Guterres, "Secretary-General's statement on the IPCC Working Group 1 Report on the Physical Science Basis of the Sixth Assessment," Aug. 2021. [Online]. Available: https://www.un.org/sg/en/content/sg/statement/2021-08-09/secretary-generals-statement-the-ipcc-working-group-1-report-the-physical-science-basis-of-the-sixth-assessment

[4] IPCC, "Summary for Policymakers," in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, Eds. Cambridge, United Kingdom and New York, NY, USA: Cambridge Univeristy Press, 2021, pp. 3–32.

[5] T. M. Lenton, D. I. Armstrong McKay, S. Loriani, J. F. Abrams, S. J. Lade, J. F. Donges, M. Milkoreit, T. Powell, S. R. Smith, C. Zimm, J. E. Buxton, E. Bailey,

L. Laybourn, A. Ghadiali, and J. G. Dyke, Eds., *The Global Tipping Points Report 2023*. Exeter, UK: University of Exeter, 2023.

[6] R. H. Cowie, P. Bouchet, and B. Fontaine, "The Sixth Mass Extinction: Fact, fiction or speculation?" *Biological Reviews*, vol. 97, no. 2, pp. 640–663, 2022.

[7] J. Rockström, J. Gupta, D. Qin, S. J. Lade, J. F. Abrams, L. S. Andersen, D. I. Armstrong McKay, X. Bai, G. Bala, S. E. Bunn, D. Ciobanu, F. DeClerck, K. Ebi, L. Gifford, C. Gordon, S. Hasan, N. Kanie, T. M. Lenton, S. Loriani, D. M. Liverman, A. Mohamed, N. Nakicenovic, D. Obura, D. Ospina, K. Prodani, C. Rammelt, B. Sakschewski, J. Scholtens, B. Stewart-Koster, T. Tharammal, D. van Vuuren, P. H. Verburg, R. Winkelmann, C. Zimm, E. M. Bennett, S. Bringezu, W. Broadgate, P. A. Green, L. Huang, L. Jacobson, C. Ndehedehe, S. Pedde, J. Rocha, M. Scheffer, L. Schulte-Uebbing, W. de Vries, C. Xiao, C. Xu, X. Xu, N. Zafra-Calvo, and X. Zhang, "Safe and just Earth system boundaries," *Nature*, vol. 619, no. 7968, pp. 102–111, Jul. 2023.

[8] M. Besson, J. Alison, K. Bjerge, T. E. Gorochowski, T. T. Høye, T. Jucker, H. M. R. Mann, and C. F. Clements, "Towards the fully automated monitoring of ecological communities," *Ecology Letters*, vol. 25, no. 12, pp. 2753–2775, 2022.

[9] V. L. Boult, "Forecast-based action for conservation," *Conservation Biology*, vol. 37, no. 3, 2023.

[10] J. Bradford, J. Weltzin, M. McCormick, J. Baron, Z. Bowen, S. Bristol, D. Carlisle, T. Crimmins, P. Cross, J. DeVivo, M. Dietze, M. Freeman, J. Goldberg, M. Hooten, L. Hsu, K. Jenni, J. Keisman, J. Kennen, K. Lee, D. Lesmes, K. Loftin, B. Miller, P. Murdoch, J. Newman, K. Prentice, I. Rangwala, J. Read, J. Sieracki, H. Sofaer, S. Thur, G. Toevs, F. Werner, C. White, T. White, and M. Wiltermuth, "Ecological forecasting—21st century science for 21st century management," U.S. Geological Survey, Open-File Report 2020–1073, 2020.

[11] B. O'Connor, S. Bojinski, C. Röösli, and M. E. Schaepman, "Monitoring global changes in biodiversity and climate essential as ecological crisis intensifies," *Ecological Informatics*, vol. 55, p. 101033, Jan. 2020.

[12] R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones, "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring," *Methods in Ecology and Evolution*, vol. 10, no. 2, pp. 169–185, 2019.

[13] L. S. M. Sugai, T. S. F. Silva, J. W. Ribeiro, Jr, and D. Llusia, "Terrestrial Passive Acoustic Monitoring: Review and Perspectives," *BioScience*, vol. 69, no. 1, pp. 15–25, Jan. 2019.

[14] S. R. P.-J. Ross, D. P. O'Connell, J. L. Deichmann, C. Desjonquères, A. Gasc, J. N. Phillips, S. S. Sethi, C. M. Wood, and Z. Burivalova, "Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions," *Functional Ecology*, vol. 37, no. 4, pp. 959–975, 2023.

[15] S. L. Dumyahn and B. C. Pijanowski, "Soundscape conservation," *Landscape Ecology*, vol. 26, no. 9, pp. 1327–1344, Nov. 2011.

[16] B. Krause, "The Niche Hypothesis: A virtual symphony of animal sounds, the origins of musical expression and the health of habitats," *Soundscape Newsletter (World Forum for Acoustic Ecology)*, Jun. 1993.

[17] Y. Seki, Ed., *Acoustic Communication in Animals: From Insect Wingbeats to Human Music (Bioacoustics Series Vol.1)*. Singapore: Springer Nature, 2023.

[18] J. W. Bradbury and S. L. Vehrencamp, *Principles of Animal Communication*, 2nd ed., ser. Animal Communication. Sunderland, MA, USA: Sinauer Associates, 2011.

[19] A. J. Oxenham, "How We Hear: The Perception and Neural Coding of Sound," *Annual Review of Psychology*, vol. 69, no. Volume 69, 2018, pp. 27–50, Jan. 2018.

[20] C. G. Scanes, "Chapter 1 - Animal Perception Including Differences With Humans," in *Animals and Human Society*, C. G. Scanes and S. R. Toukhsati, Eds. Academic Press, Jan. 2018, pp. 1–11.

[21] L. R. V. Gilmour, I. Bray, C. Alford, and P. R. Lintott, "Natural soundscapes enhance mood recovery amid anthropogenic noise pollution," *PLOS ONE*, vol. 19, no. 11, p. e0311487, Nov. 2024.

[22] J.-Y. Lin, W.-J. Cheng, C.-F. Wu, and T.-Y. Chang, "Associations of road traffic noise and its frequency spectrum with prevalent depression in Taichung, Taiwan," *Frontiers in Public Health*, vol. 11, Jan. 2023.

[23] E. Ratcliffe, "Sound and Soundscape in Restorative Natural Environments: A Narrative Literature Review," *Frontiers in Psychology*, vol. 12, Apr. 2021.

[24] T. D. Siegel, W. J. Cooper, R. E. Forkner, W. F. Laurance, J. Luís Camargo, and D. Luther, "Forest fragmentation effects on mutualistic interactions: Frugivorous birds and fruiting trees," *Oikos*, vol. 2024, no. 10, p. e10383, 2024.

[25] E. Znidersic and D. M. Watson, "Acoustic restoration: Using soundscapes to benchmark and fast-track recovery of ecological communities," *Ecology Letters*, vol. 25, no. 7, pp. 1597–1603, 2022.

[26] Z. Burivalova, T. M. Maeda, Purnomo, Y. Rayadin, T. Boucher, P. Choksi, P. Roe, A. Truskinger, and E. T. Game, "Loss of temporal structure of tropical soundscapes with intensifying land use in Borneo," *Science of The Total Environment*, vol. 852, p. 158268, Dec. 2022.

[27] J. Lawson, A. Whitworth, and C. Banks-Leite, "Soundscapes show disruption across the diel cycle in human modified tropical landscapes," *Ecological Indicators*, vol. 144, p. 109413, Nov. 2022.

[28] P. Somervuo, T. Roslin, B. L. Fisher, B. Hardwick, D. Kerdraon, D. Raharinjanahary, E. T. Rajoelison, P. Lauha, L. Griem, P. Lehikoinen, P. Niittynen, E. Piirainen, M. Lumme, V.-M. Riihikoski, O. Acevedo-Charry, S. A. Adalsteinsson, M. Ahmad, S. Alcobia, J. Aldará, N. R. Andrew, S. Anslan, A. Antonelli, J. S. Arena, S. Arroyo Almeida, I. Aster, H. Autto, A. Aviles Gamboa, J. Baixeras, M. Baldauf, R. Balestrieri, G. G. Banelyte, A. Barrett, P. Beja, T. O. Berg, B. Bergerot, E. G. Biro, P. G. Blendinger, L. Bollache, M. Bou Dagher Kharrat, S. Boyer, E. Bridell, M. Brotherson, L. R. Brown, H. L. Buckley, E. Buscardo, N. Buthelezi, L. Cagnolo, A. Calvente, G. Capobianco, L. Carreón-Palau, S. Carriere, B. S. Case, J. Chang, J. M. Chaparro, C.-L. Chen, C. Chicoine, M. Christensson, F. Collado Rosique, W. Colom Montero, R. d. S. da Fonseca, L. P. Da Silva, A. Dal Molin, T. Dallas, M. C. de Francesco, J. A. Del Ángel-Rodríguez, R. Díaz-Delgado, T. Dirnböck,

I. Djukic, P. Dladla, J. Domínguez Masciale, T. Dorigo, E. Douwes, T. Ekrem, H. Enderskog, C. Erefur, M. Fahad, M. Falahati-Anbaran, A. Farrell, G. Ferland, E. Ferrari, A. Figueiredo, F. Forero, I. Freiberga, A. Frosch-Radivo, L. A. Ganchozo Intriago, L. Garzoli, P. Giacomotti, A. T. Gianuca, O. Gilg, V. Gilg, F. Gísladóttir, R. Glowacki, B. Gottsberger, J. Gregoire, E. Groner, P. Guedes, A. M. Guile, P. Haase, F. Hadi, M. Haidegger, L. J. Hansen, L. H. Hansen, R. Harrop, H. Havnås, D. Herrera Báez, C. C. Y. Ho, D. Hohenbühel, M. Houska Tahadlova, J. Hänninen, L. Höglund, K. Í Haraldsstovu, E. Imbeau, J. Inkinen, M. I. Ishihara, A. C. Jackson, G. Jansson, R. Jha, G. Kager, R. Kahale, O. Kalttopää, E. W. Karai, D. Karlsson, A. Kaus-Thiel, A. Khan, Q. Khan, K. Kimoto, S. C. Kipngetich, C. Klante, L. Klemedtsson, M. Klinth, J. Koskinen, M. Kotakorpi, A.-K. Kreiling, I. Krisai-Greilhuber, E. Kristensen, S. König, S. Langenheder, K. Laurila, P. Le Gouar, N. Lecomte, E. Lecomte, P. M. Leitman, J. L. León-Cortés, D. Li, J. Loehr, C. Lopez-Vaamonde, M. Makari, G. G. Mangini, M. Maroschek, V. A. Mata, S. Matsuoka, T. Mazzafera, P. G. McDonald, L. Meinert, M. Meléndez-González, A. M. Mendoza-Henao, S. Moreau, J. Moreau, J. Mosbacher, E. Moyer, A. Mrazova, S. Mteshane, N. W. Mungai, G. Muñoz Herraiz, A. Murillo-Vázquez, S. Musazzi, M. Mutanen, J. Müller, R. Navarro Canales, M. Ndlovu, A. Nicolai, A. Niessner, J. Nordén, P. Nowak, E. O'Connell, A. Orru, T. Pagnon, Y. N. Pantoja-Diaz, M. Pentinsaari, S. Pilloni, A. Pinder, T. A. Pinheiro, S. Põlme, L. L. Powell, G. Pröll, P. Pulido-Santacruz, E. Queralt, M. T. Quilantang, K. Quinlan, R. Ramirez, J. Rankinen, M. D. V. Rasino, R. Rebelo, W. Remmers, F. Retz, E. Reyes, G. Rivas Torres, H. M. K. Rogers, I. T. Rosário, S. Rosário Da Rosàrio da Costa, T. Rütting, J. Sahlstén, C. Saliba, T. Salmirinne, K. Sam, D. Santos, M. Santos-Reis, M. Sawan, B. Schattanek-Wiesmair, P. Schiestl-Aalto, N. M. Schmidt, S. Seibold, R. Seidl, L. Seifert, M. Sithole, E. Sivault, J. Smart, I. Smerczyński, A. Soda, R. S. Sousa-Lima, A. Stanisci, M. C. Stanley, D. Steenkamp, E. Stengel, S. Stoll, W. M. Strauss, E. Stur, M. Sujala, J. Sundell, J. Svavarsdóttir, L. Tedersoo, S. Tepsa, M. T. Tikochinsky, E.-P. Tuominen, S. Tweraser, C. Ulloa Espinosa, J. Uusitalo, M. Vallinmäki, F. Vannier, A. Varela, E. Vatka, S. Veikkolainen, K. Vernes, P. C. Watts, P. Weslien, C. Wirth, J. H. Wisniewski, A. B. Young, R. Övergaard, and O. Ovaskainen, "Human contributions to global

soundscapes are less predictable than the acoustic rhythms of wildlife," *Nature Ecology & Evolution*, pp. 1–14, Jul. 2025.

[29] N. Bustamante and Á. Garitano-Zavala, "Natural Patterns in the Dawn and Dusk Choruses of a Neotropical Songbird in Relation to an Urban Sound Environment," *Animals*, vol. 14, no. 4, p. 646, Jan. 2024.

[30] R. Carson, *Silent Spring*. Boston, MA, USA: Houghton Mifflin Company, 1962.

[31] B. C. Pijanowski, L. J. Villanueva-Rivera, S. L. Dumyahn, A. Farina, B. L. Krause, B. M. Napoletano, S. H. Gage, and N. Pieretti, "Soundscape Ecology: The Science of Sound in the Landscape," *BioScience*, vol. 61, no. 3, pp. 203–216, Mar. 2011.

[32] J. Sueur and A. Farina, "Ecoacoustics: The Ecological Investigation and Interpretation of Environmental Sound," *Biosemiotics*, vol. 8, no. 3, pp. 493–502, Dec. 2015.

[33] N. Fletcher, "Animal Bioacoustics," in *Springer Handbook of Acoustics*, T. Rossing, Ed. Springer, New York, NY, 2007, pp. 785–804.

[34] R. M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, VT: Destiny Books, 1994.

[35] E. S. Morton, "Ecological Sources of Selection on Avian Sounds," *The American Naturalist*, Jan. 1975.

[36] I. Alcocer, H. Lima, L. S. M. Sugai, and D. Llusia, "Acoustic indices as proxies for biodiversity: A meta-analysis," *Biological Reviews*, vol. 97, no. 6, pp. 2209–2236, 2022.

[37] T. Bradfer-Lawrence, C. Desjonqueres, A. Eldridge, A. Johnston, and O. Metcalf, "Using acoustic indices in ecology: Guidance on study design, analyses and interpretation," *Methods in Ecology and Evolution*, vol. 14, no. 9, pp. 2192–2204, 2023.

[38] A. P. Hill, P. Prince, E. Piña Covarrubias, C. P. Doncaster, J. L. Snaddon, and A. Rogers, "AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment," *Methods in Ecology and Evolution*, vol. 9, no. 5, pp. 1199–1211, 2018.

[39] R. Koch, M. Raymond, P. Wrege, and H. Klinck, "SWIFT: A small, low-cost acoustic recorder for terrestrial wildlife monitoring applications," in *North American Ornithological Conference. Washington, DC*, 2016, p. 619.

[40] S. S. Sethi, R. M. Ewers, N. S. Jones, C. D. L. Orme, and L. Picinali, "Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device," *Methods in Ecology and Evolution*, vol. 9, no. 12, pp. 2383–2387, Dec. 2018.

[41] S. S. Sethi, R. M. Ewers, N. S. Jones, A. Signorelli, L. Picinali, and C. D. L. Orme, "SAFE Acoustics: An open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo," *Methods in Ecology and Evolution*, vol. 11, no. 10, pp. 1182–1185, Oct. 2020.

[42] T. A. Rhinehart, L. M. Chronister, T. Devlin, and J. Kitzes, "Acoustic localization of terrestrial wildlife: Current practices and future opportunities," *Ecology and Evolution*, vol. 10, no. 13, pp. 6794–6818, 2020.

[43] W. A. Watkins and W. E. Schevill, "Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array," *Deep Sea Research and Oceanographic Abstracts*, vol. 19, no. 10, pp. 691–706, Oct. 1972.

[44] D. J. Mennill, M. Battiston, D. R. Wilson, J. R. Foote, and S. M. Doucet, "Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour," *Methods in Ecology and Evolution*, vol. 3, no. 4, pp. 704–712, 2012.

[45] P. H. Wrege, E. D. Rowland, S. Keen, and Y. Shiu, "Acoustic monitoring for conservation in tropical forests: Examples from forest elephants," *Methods in Ecology and Evolution*, vol. 8, no. 10, pp. 1292–1301, 2017.

[46] D. T. Blumstein, D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, S. F. Hanser, B. McCowan, A. M. Ali, and A. N. G. Kirschel, "Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus," *Journal of Applied Ecology*, vol. 48, no. 3, pp. 758–767, 2011.

[47] M. E. Lockwood, D. L. Jones, R. C. Bilger, C. R. Lansing, W. D. O'Brien, Jr., B. C. Wheeler, and A. S. Feng, "Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms," *The Journal of the Acoustical Society of America*, vol. 115, no. 1, pp. 379–391, Dec. 2003.

[48] M. Allen, L. Girod, R. Newton, S. Madden, D. T. Blumstein, and D. Estrin, "VoxNet: An Interactive, Rapidly-Deployable Acoustic Monitoring Platform," in *Proceedings of the 7th International Conference on Information Processing in Sensor Networks*, ser. IPSN '08. USA: IEEE Computer Society, Apr. 2008, pp. 371–382.

[49] B. E. Heath, R. Suzuki, N. P. Le Penru, J. Skinner, C. D. L. Orme, R. M. Ewers, S. S. Sethi, and L. Picinali, "Spatial ecosystem monitoring with a Multichannel Acoustic Autonomous Recording Unit (MAARU)," *Methods in Ecology and Evolution*, vol. 15, no. 9, pp. 1568–1579, 2024.

[50] P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, "A survey of sound source localization with deep learning methods," *The Journal of the Acoustical Society of America*, vol. 152, no. 1, pp. 107–151, Jul. 2022.

[51] L. Bouffaut, K. Taweesintananon, H. J. Kriesell, R. A. Rørstadbotnen, J. R. Potter, M. Landrø, S. E. Johansen, J. K. Brenne, A. Haukanes, O. Schjelderup, and F. Storvik, "Eavesdropping at the Speed of Light: Distributed Acoustic Sensing of Baleen Whales in the Arctic," *Frontiers in Marine Science*, vol. 9, Jul. 2022.

[52] E. R. Snyder, A. Solsona-Berga, S. Baumann-Pickering, K. E. Frasier, S. M. Wiggins, and J. A. Hildebrand, "Where's Whaledo: A software toolkit for array localization of animal vocalizations," *PLOS Computational Biology*, vol. 20, no. 5, p. e1011456, May 2024.

[53] R. Suzuki, S. Matsubayashi, R. W. Hedley, K. Nakadai, and H. G. Okuno, "HARKBird: Exploring Acoustic Interactions in Bird Communities Using a Microphone Array," *Journal of Robotics and Mechatronics*, vol. 29, no. 1, pp. 213–223, Feb. 2017.

[54] S. Sumitani, R. Suzuki, N. Chiba, S. Matsubayashi, T. Arita, K. Nakadai, and H. G. Okuno, "An Integrated Framework for Field Recording, Localization, Classification

and Annotation of Birdsongs Using Robot Audition Techniques — Harkbird 2.0," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, Unilted Kingdom, May 2019, pp. 8246–8250.

[55] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[56] G. Grimm, J. Luberadzka, and V. Hohmann, "A Toolbox for Rendering Virtual Acoustic Environments in the Context of Audiology," *Acta Acustica united with Acustica*, vol. 105, no. 3, pp. 566–578, May 2019.

[57] L. S. R. Simon, N. Dillier, and H. Wüthrich, "Comparison of 3D Audio Reproduction Methods Using Hearing Devices," *J. Audio Eng. Soc*, vol. 68, no. 12, pp. 899–909, 2020.

[58] A. De Rosa, I. Castro, and S. Marsland, "The acoustic playback technique in avian fieldwork contexts: A systematic review and recommendations for best practice," *Ibis*, vol. 164, no. 2, pp. 371–387, 2022.

[59] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, "Tipping elements in the Earth's climate system," *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 1786–1793, Feb. 2008.

[60] V. Dakos, C. A. Boulton, J. E. Buxton, J. F. Abrams, B. Arellano-Nava, D. I. Armstrong McKay, S. Bathiany, L. Blaschke, N. Boers, D. Dylewsky, C. López-Martínez, I. Parry, P. Ritchie, B. van der Bolt, L. van der Laan, E. Weinans, and S. Kéfi, "Tipping point detection and early warnings in climate, ecological, and human systems," *Earth System Dynamics*, vol. 15, no. 4, pp. 1117–1135, Aug. 2024.

[61] T. M. Lenton, J. Rockström, O. Gaffney, S. Rahmstorf, K. Richardson, W. Steffen, and H. J. Schellnhuber, "Climate tipping points — too risky to bet against," *Nature*, vol. 575, no. 7784, pp. 592–595, Nov. 2019.

[62] C. A. Boulton, T. M. Lenton, and N. Boers, "Pronounced loss of Amazon rainforest resilience since the early 2000s," *Nature Climate Change*, pp. 1–8, Mar. 2022.

[63] B. M. Flores, E. Montoya, B. Sakschewski, N. Nascimento, A. Staal, R. A. Betts, C. Levis, D. M. Lapola, A. Esquível-Muelbert, C. Jakovac, C. A. Nobre, R. S.

Oliveira, L. S. Borma, D. Nian, N. Boers, S. B. Hecht, H. ter Steege, J. Arieira, I. L. Lucas, E. Berenguer, J. A. Marengo, L. V. Gatti, C. R. C. Mattos, and M. Hirota, "Critical transitions in the Amazon forest system," *Nature*, vol. 626, no. 7999, pp. 555–564, Feb. 2024.

[64] M. Hirota, M. Holmgren, E. H. Van Nes, and M. Scheffer, "Global Resilience of Tropical Forest and Savanna to Critical Transitions," *Science*, vol. 334, no. 6053, pp. 232–235, Oct. 2011.

[65] A. Staal, O. A. Tuinenburg, J. H. C. Bosmans, M. Holmgren, E. H. van Nes, M. Scheffer, D. C. Zemp, and S. C. Dekker, "Forest-rainfall cascades buffer against drought across the Amazon," *Nature Climate Change*, vol. 8, no. 6, pp. 539–543, Jun. 2018.

[66] D. C. Zemp, C.-F. Schleussner, H. M. J. Barbosa, M. Hirota, V. Montade, G. Sampaio, A. Staal, L. Wang-Erlandsson, and A. Rammig, "Self-amplified Amazon forest loss due to vegetation-atmosphere feedbacks," *Nature Communications*, vol. 8, no. 1, p. 14681, Mar. 2017.

[67] P. Bunyard and R. De Laet, *Cooling the Climate: How to Revive the Biosphere and Cool the Earth within 20 Years.* Ethics International Press Limited, UK, 2024.

[68] M. Scheffer, *Critical Transitions in Nature and Society.* Princeton University Press, 2009.

[69] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 3rd ed. Boca Raton, FL: CRC Press, 2024.

[70] C. Folke, S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, L. Gunderson, and C. Holling, "Regime Shifts, Resilience, and Biodiversity in Ecosystem Management," *Annual Review of Ecology, Evolution, and Systematics*, vol. 35, no. 1, pp. 557–581, 2004.

[71] S. R. Carpenter, J. J. Cole, M. L. Pace, R. Batt, W. A. Brock, T. Cline, J. Coloso, J. R. Hodgson, J. F. Kitchell, D. A. Seekell, L. Smith, and B. Weidel, "Early Warnings of Regime Shifts: A Whole-Ecosystem Experiment," *Science*, vol. 332, no. 6033, pp. 1079–1082, May 2011.

[72] V. Dakos, S. R. Carpenter, E. H. van Nes, and M. Scheffer, "Resilience indicators: Prospects and limitations for early warnings of regime shifts," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1659, p. 20130263, Jan. 2015.

[73] M. Sánchez-Pinillos, V. Dakos, and S. Kéfi, "Ecological dynamic regimes: A key concept for assessing ecological resilience," *Biological Conservation*, vol. 289, p. 110409, Jan. 2024.

[74] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara, "Early-warning signals for critical transitions," *Nature*, vol. 461, no. 7260, pp. 53–59, Sep. 2009.

[75] M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*, 1st ed. Boston: Little, Brown, 2000.

[76] E. H. van Nes, B. M. S. Arani, A. Staal, B. van der Bolt, B. M. Flores, S. Bathiany, and M. Scheffer, "What Do You Mean, 'Tipping Point'?" *Trends in Ecology & Evolution*, vol. 31, no. 12, pp. 902–904, Dec. 2016.

[77] W. Liu, S.-P. Xie, Z. Liu, and J. Zhu, "Overlooked possibility of a collapsed Atlantic Meridional Overturning Circulation in warming climate," *Science Advances*, vol. 3, no. 1, p. e1601666, Jan. 2017.

[78] M. Scheffer, S. R. Carpenter, J. A. Foley, C. Folke, and B. Walker, "Catastrophic shifts in ecosystems," *Nature*, vol. 413, no. 6856, pp. 591–596, Oct. 2001.

[79] T. M. Lenton, "Environmental Tipping Points," *Annual Review of Environment and Resources*, vol. 38, no. Volume 38, 2013, pp. 1–29, Oct. 2013.

[80] S. Kéfi, V. Guttal, W. A. Brock, S. R. Carpenter, A. M. Ellison, V. N. Livina, D. A. Seekell, M. Scheffer, E. H. van Nes, and V. Dakos, "Early Warning Signals of Ecological Transitions: Methods for Spatial Patterns," *PLoS ONE*, vol. 9, no. 3, p. e92097, Mar. 2014.

[81] M. Scheffer, S. R. Carpenter, V. Dakos, and E. H. van Nes, "Generic Indicators of Ecological Resilience: Inferring the Chance of a Critical Transition," *Annual Review of Ecology, Evolution, and Systematics*, vol. 46, no. 1, pp. 145–167, 2015.

[82] J. C. Rocha, G. Peterson, Ö. Bodin, and S. Levin, "Cascading regime shifts within and across scales," *Science*, vol. 362, no. 6421, pp. 1379–1383, Dec. 2018.

[83] N. Boers and M. Rypdal, "Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point," *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, p. e2024192118, May 2021.

[84] T. M. Lenton, J. E. Buxton, D. I. Armstrong McKay, J. F. Abrams, C. A. Boulton, K. Lees, T. W. R. Powell, N. Boers, A. M. Cunliffe, and V. Dakos, "A resilience sensing system for the biosphere," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 377, no. 1857, p. 20210383, Jun. 2022.

[85] T. M. Lenton, J. F. Abrams, A. Bartsch, S. Bathiany, C. A. Boulton, J. E. Buxton, A. Conversi, A. M. Cunliffe, S. Hebden, T. Lavergne, B. Poulter, A. Shepherd, T. Smith, D. Swingedouw, R. Winkelmann, and N. Boers, "Remotely sensing potential climate change tipping points across scales," *Nature Communications*, vol. 15, no. 1, p. 343, Jan. 2024.

[86] R. M. van Westen, M. Kliphuis, and H. A. Dijkstra, "Physics-based early warning signal shows that AMOC is on tipping course," *Science Advances*, Feb. 2024.

[87] P. Ashwin, S. Wieczorek, R. Vitolo, and P. Cox, "Tipping points in open systems: Bifurcation, noise-induced and rate-dependent examples in the climate system," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1962, pp. 1166–1184, Mar. 2012.

[88] Y. A. Kuznetsov, *Elements of Applied Bifurcation Theory*, ser. Applied Mathematical Sciences. Cham: Springer International Publishing, 2023, vol. 112.

[89] C. Wissel, "A universal law of the characteristic return time near thresholds," *Oecologia*, vol. 65, no. 1, pp. 101–107, Dec. 1984.

[90] C. S. Holling, "Resilience and Stability of Ecological Systems," *Annual Review of Ecology and Systematics*, vol. 4, no. 1, pp. 1–23, 1973.

[91] ——, *Engineering Resilience versus Ecological Resilience*. Washington, DC: The National Academies Press, 1996, pp. 31–44.

[92] S. Carpenter, B. Walker, J. M. Anderies, and N. Abel, "From Metaphor to Measurement: Resilience of What to What?" *Ecosystems*, vol. 4, no. 8, pp. 765–781, Dec. 2001.

[93] D. Hodgson, J. L. McDonald, and D. J. Hosken, "What do you mean, 'resilient'?" *Trends in Ecology & Evolution*, vol. 30, no. 9, pp. 503–506, Sep. 2015.

[94] B. Walker, "Resilience: What it is and is not," *Ecology and Society*, vol. 25, no. 2, May 2020.

[95] V. Dakos and S. Kéfi, "Ecological resilience: What to measure and how," *Environmental Research Letters*, vol. 17, no. 4, p. 043003, Mar. 2022.

[96] A. J. Lotka, *Elements of Physical Biology*. Baltimore, MD: Williams & Wilkins company, 1925.

[97] R. C. Lewontin, "The Meaning of Stability," in *Diversity and Stability in Ecological Systems: A Report of Symposium Held May 26-28, 1969*, ser. Brookhaven Symposia in Biology, G. M. Woodwell and H. H. Smith, Eds. Upton, NY: Biology Dept., Brookhaven National Laboratory, 1969, no. 22, pp. 13–24.

[98] R. M. May, "Thresholds and breakpoints in ecosystems with a multiplicity of stable states," *Nature*, vol. 269, no. 5628, pp. 471–477, Oct. 1977.

[99] I. Noy-Meir, "Stability of Grazing Systems: An Application of Predator-Prey Graphs," *Journal of Ecology*, vol. 63, no. 2, pp. 459–481, 1975.

[100] M. L. Rosenzweig and R. H. MacArthur, "Graphical Representation and Stability Conditions of Predator-Prey Interactions," *The American Naturalist*, 1963.

[101] R. Kubo, "The fluctuation-dissipation theorem," *Reports on Progress in Physics*, vol. 29, no. 1, p. 255, Jan. 1966.

[102] A. R. Ives, "Measuring Resilience in Stochastic Systems," *Ecological Monographs*, vol. 65, no. 2, pp. 217–233, 1995.

[103] H. Held and T. Kleinen, "Detection of climate system bifurcations by degenerate fingerprinting," *Geophysical Research Letters*, vol. 31, no. 23, 2004.

[104] S. R. Carpenter and W. A. Brock, "Rising variance: A leading indicator of ecological transition," *Ecology Letters*, vol. 9, no. 3, pp. 311–318, 2006.

[105] C. F. Clements and A. Ozgul, "Indicators of transitions in biological systems," *Ecology Letters*, vol. 21, no. 6, pp. 905–919, Jun. 2018.

[106] C. F. Clements, J. M. Drake, J. I. Griffiths, and A. Ozgul, "Factors Influencing the Detectability of Early Warning Signals of Population Collapse," *The American Naturalist*, vol. 186, no. 1, pp. 50–58, Jul. 2015.

[107] V. Dakos, M. Scheffer, E. H. van Nes, V. Brovkin, V. Petoukhov, and H. Held, "Slowing down as an early warning signal for abrupt climate change," *Proceedings of the National Academy of Sciences*, vol. 105, no. 38, pp. 14 308–14 312, Sep. 2008.

[108] V. N. Livina, F. Kwasniok, and T. M. Lenton, "Potential analysis reveals changing number of climate states during the last 60 kyr," *Climate of the Past*, vol. 6, no. 1, pp. 77–82, Feb. 2010.

[109] V. Guttal and C. Jayaprakash, "Spatial variance and spatial skewness: Leading indicators of regime shifts in spatial ecological systems," *Theoretical Ecology*, vol. 2, no. 1, pp. 3–12, Mar. 2009.

[110] C. Ciemer, N. Boers, M. Hirota, J. Kurths, F. Müller-Hansen, R. S. Oliveira, and R. Winkelmann, "Higher resilience to climatic disturbances in tropical vegetation exposed to more variable rainfall," *Nature Geoscience*, vol. 12, no. 3, pp. 174–179, Mar. 2019.

[111] F. de Oliveira Roque, J. F. S. Menezes, T. Northfield, J. M. Ochoa-Quintero, M. J. Campbell, and W. F. Laurance, "Warning signals of biodiversity collapse across gradients of tropical forest loss," *Scientific Reports*, vol. 8, no. 1, p. 1622, Jan. 2018.

[112] S. Eby, A. Agrawal, S. Majumder, A. P. Dobson, and V. Guttal, "Alternative stable states and spatial indicators of critical slowing down along a spatial gradient in a savanna ecosystem," *Global Ecology and Biogeography*, vol. 26, no. 6, pp. 638–649, 2017.

[113] J. Verbesselt, N. Umlauf, M. Hirota, M. Holmgren, E. H. Van Nes, M. Herold, A. Zeileis, and M. Scheffer, "Remotely sensed resilience of tropical forests," *Nature Climate Change*, vol. 6, no. 11, pp. 1028–1031, Nov. 2016.

[114] S. S. Farley, A. Dawson, S. J. Goring, and J. W. Williams, "Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions," *BioScience*, vol. 68, no. 8, pp. 563–576, Aug. 2018.

[115] S. L. Pimm, S. Alibhai, R. Bergl, A. Dehgan, C. Giri, Z. Jewell, L. Joppa, R. Kays, and S. Loarie, "Emerging Technologies to Conserve Biodiversity," *Trends in Ecology & Evolution*, vol. 30, no. 11, pp. 685–696, Nov. 2015.

[116] K. F. A. Darras, R. A. Rountree, S. L. Van Wilgenburg, A. F. Cord, F. Pitz, Y. Chen, L. Dong, A. Rocquencourt, C. Desjonquères, P. M. Diaz, T.-H. Lin, T. Turco, L. Emmerson, T. Bradfer-Lawrence, A. Gasc, S. Marley, M. Salton, L. Schillé, P. J. Wensveen, S.-H. Wu, A. C. Acero-Murcia, O. Acevedo-Charry, M. Adam, J. Aguzzi, I. Akoglu, M. C. P. Amorim, M. Anders, M. André, A. Antonelli, L. A. Do Nascimento, G. Appel, S. Archer, C. Astaras, A. Atemasov, J. Atkinson, J. Attia, E. Baltag, L. Barbaro, F. Basan, C. Batist, J. E. Baumgarten, J. T. Bayle Sempere, K. Bellisario, A. B. David, O. Berger-Tal, F. Bertucci, M. G. Betts, I. S. Bhalla, T. Bicudo, M. Bolgan, S. Bombaci, G. Bota, M. Boullhesen, R. A. Briers, S. Buchan, M. Budka, K. Burchard, G. Buscaino, A. Calvente, M. Campos-Cerqueira, M. I. C. Gonçalves, M. Ceraulo, M. Cerezo-Araujo, G. Cerwén, A. A. Chaskda, M. Chistopolova, C. W. Clark, K. D. Cox, B. Cretois, C. Czarnecki, L. P. da Silva, W. da Silva, L. H. De Clippele, D. de la Haye, A. S. de Oliveira Tissiani, D. de Zwaan, M. E. Degano, J. Deichmann, J. del Rio, C. Devenish, R. Díaz-Delgado, P. Diniz, D. D. Oliveira-Júnior, T. Dorigo, S. Dröge, M. Duarte, A. Duarte, K. Dunleavy, R. Dziak, S. Elise, H. Enari, H. S. Enari, F. Erbs, B. K. Eriksson, P. Ertör-Akyazi, N. C. Ferrari, L. Ferreira, A. B. Fleishman, P. J. Fonseca, B. Freitas, N. R. Friedman, J. S. P. Froidevaux, S. Gogoleva, C. Gonzaga, J. M. G. Correa, E. Goodale, B. Gottesman, I. Grass, J. Greenhalgh, J. Gregoire, S. Haché, J. Hagge, W. Halliday, A. Hammer, T. Hanf-Dressler, S. Haupert, S. Haver, B. Heath, D. Hending, J. Hernandez-Blanco, D. Higgs, T. Hiller, J. C.-C. Huang, K. L. Hutchinson, C. Hyacinthe, C. Ieronymidou, I. A. Iniunam,

J. Jackson, A. Jacot, O. Jahn, F. Juanes, K. S. J. Kanes, E. Kenchington, S. Kepfer-Rojas, J. Kitzes, T. Kusuminda, Y. Lehnardt, J. Lei, P. Leitman, J. León, D. Li, C. S. Lima-Santos, K. J. Lloyd, A. Looby, A. López-Baucells, D. López-Bosch, T. Louth-Robins, T. Maeda, F. Malige, C. Mammides, G. Marcacci, M. Markolf, M. I. Marques, C. W. Martin, D. A. Martin, K. Martin, E. McArthur, M. McKown, L. J. T. McLeod, V. Médoc, O. Metcalf, C. F. J. Meyer, G. Mikusinski, B. Miller, J. Monteiro, T. A. Mooney, S. Moreira, L. S. M. Sugai, D. Morris, S. Müller, S. Muñoz-Duque, K. A. Murchy, I. Nagelkerken, M. Mas, R. Nouioua, C. Ocampo-Ariza, J. D. Olden, S. Oppel, A. N. Osiecka, E. Papale, M. Parsons, M. Pashkevich, J. Patris, J. P. Marques, C. Pérez-Granados, L. Piatti, M. Pichorim, M. K. Pine, T. Pinheiro, J.-N. Pradervand, J. Quinn, B. Quintella, C. Radford, X. Raick, A. Rainho, E. Ramalho, V. Ramesh, S. Rétaux, L. K. Reynolds, K. Riede, T. Rimmer, N. Ríos, R. Rocha, L. Rocha, P. Roe, S. R. P.-J. Ross, C. M. Rosten, J. Ryan, C. Salustio-Gomes, F. I. P. Samarra, P. Samartzis, J. Santos, T. Sattler, K. Scharffenberg, R. P. Schoeman, K.-L. Schuchmann, E. Sebastián-González, S. Seibold, S. Sethi, F. W. Shabangu, T. Shaw, X. Shen, D. Singer, A. Širović, M. Slater, B. Spriel, J. Stanley, J. Sueur, V. da Cunha Tavares, K. Thomisch, S. Thorn, J. Tong, L. Torrent, J. Traba, J. A. Tremblay, L. Trevelin, S. Tseng, M.-N. Tuanmu, M. Valverde, B. Vernasco, M. Vieira, R. V. da Paz, M. Ward, M. Watson, M. J. Weldy, J. Wiel, J. Willie, H. Wood, J. Xu, W. Zhou, S. Li, R. Sousa-Lima, and T. C. Wanger, "Worldwide Soundscapes: A Synthesis of Passive Acoustic Monitoring Across Realms," *Global Ecology and Biogeography*, vol. 34, no. 5, p. e70021, 2025.

[117] S. M. Van Parijs, C. W. Clark, R. S. Sousa-Lima, S. E. Parks, S. Rankin, D. Risch, and I. C. V. Opzeeland, "Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales," *Marine Ecology Progress Series*, vol. 395, pp. 21–36, Dec. 2009.

[118] A. Rice, A. Širović, J. S. Trickey, A. J. Debich, R. S. Gottlieb, S. M. Wiggins, J. A. Hildebrand, and S. Baumann-Pickering, "Cetacean occurrence in the Gulf of Alaska from long-term passive acoustic monitoring," *Marine Biology*, vol. 168, no. 5, p. 72, Apr. 2021.

[119] G. J. Measey, B. C. Stevenson, T. Scott, R. Altwegg, and D. L. Borchers, "Counting chirps: Acoustic monitoring of cryptic frogs," *Journal of Applied Ecology*, vol. 54, no. 3, pp. 894–902, 2017.

[120] A. B. Kohlberg, C. R. Myers, and L. L. Figueroa, "From buzzes to bytes: A systematic review of automated bioacoustics models used to detect, classify and monitor insects," *Journal of Applied Ecology*, vol. 61, no. 6, pp. 1199–1211, 2024.

[121] L. B. Symes, S. Madhusudhana, S. J. Martinson, C. E. Kernan, K. B. Hodge, D. P. Salisbury, H. Klinck, and H. ter Hofstede, "Estimation of katydid calling activity from soundscape recordings," *Journal of Orthoptera Research*, vol. 31, no. 2, pp. 173–180, Oct. 2022.

[122] D. J. Clink and H. Klinck, "Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring," *Methods in Ecology and Evolution*, vol. 12, no. 2, pp. 328–341, 2021.

[123] S. S. Sethi, N. S. Jones, B. D. Fulcher, L. Picinali, D. J. Clink, H. Klinck, C. D. L. Orme, P. H. Wrege, and R. M. Ewers, "Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set," *Proceedings of the National Academy of Sciences*, vol. 117, no. 29, pp. 17 049–17 055, Jul. 2020.

[124] S. R. P. J. Ross, N. R. Friedman, K. L. Dudley, M. Yoshimura, T. Yoshida, and E. P. Economo, "Listening to ecosystems: Data-rich acoustic monitoring through landscape-scale sensor networks," *Ecological Research*, vol. 33, no. 1, pp. 135–147, 2018.

[125] C. M. Wood, V. D. Popescu, H. Klinck, J. J. Keane, R. J. Gutiérrez, S. C. Sawyer, and M. Z. Peery, "Detecting small changes in populations at landscape scales: A bioacoustic site-occupancy framework," *Ecological Indicators*, vol. 98, pp. 492–507, Mar. 2019.

[126] S. S. Sethi, F. Fossøy, B. Cretois, and C. M. Rosten, *Management Relevant Applications of Acoustic Monitoring for Norwegian Nature – The Sound of Norway.* Norsk institutt for naturforskning (NINA), 2021.

[127] P. Roe, P. Eichinski, R. A. Fuller, P. G. McDonald, L. Schwarzkopf, M. Towsey, A. Truskinger, D. Tucker, and D. M. Watson, "The Australian Acoustic Observatory," *Methods in Ecology and Evolution*, vol. 12, no. 10, pp. 1802–1808, 2021.

[128] J. L. Deichmann, A. Hernández-Serna, J. A. Delgado C., M. Campos-Cerqueira, and T. M. Aide, "Soundscape analysis and acoustic monitoring document impacts of natural gas exploration on biodiversity in a tropical forest," *Ecological Indicators*, vol. 74, pp. 39–48, Mar. 2017.

[129] Z. Burivalova, M. Towsey, T. Boucher, A. Truskinger, C. Apelis, P. Roe, and E. T. Game, "Using soundscapes to detect variable degrees of human influence on tropical forests in Papua New Guinea," *Conservation Biology*, vol. 32, no. 1, pp. 205–215, 2018.

[130] A. Gasc, J. Anso, J. Sueur, H. Jourdan, and L. Desutter-Grandcolas, "Cricket calling communities as an indicator of the invasive ant Wasmannia auropunctata in an insular biodiversity hotspot," *Biological Invasions*, vol. 20, no. 5, pp. 1099–1111, May 2018.

[131] C. M. Wood, J. Socolar, S. Kahl, M. Z. Peery, P. Chaon, K. Kelly, R. A. Koch, S. C. Sawyer, and H. Klinck, "A scalable and transferable approach to combining emerging conservation technologies to identify biodiversity change after large disturbances," *Journal of Applied Ecology*, vol. 61, no. 4, pp. 797–808, 2024.

[132] B. L. Gottesman, J. C. Olson, S. Yang, O. Acevedo-Charry, D. Francomano, F. A. Martinez, R. S. Appeldoorn, D. M. Mason, E. Weil, and B. C. Pijanowski, "What does resilience sound like? Coral reef and dry forest acoustic communities respond differently to Hurricane Maria," *Ecological Indicators*, vol. 126, p. 107635, Jul. 2021.

[133] S. R. P.-J. Ross, N. R. Friedman, K. L. Dudley, T. Yoshida, M. Yoshimura, E. P. Economo, D. W. Armitage, and I. Donohue, "Divergent ecological responses to typhoon disturbance revealed via landscape-scale acoustic monitoring," *Global Change Biology*, vol. 30, no. 1, p. e17067, 2024.

[134] C. Boettiger, N. Ross, and A. Hastings, "Early warning signals: The charted and uncharted territories," *Theoretical Ecology*, vol. 6, no. 3, pp. 255–264, Aug. 2013.

[135] S. Kéfi, V. Dakos, M. Scheffer, E. H. Van Nes, and M. Rietkerk, "Early warning signals also precede non-catastrophic transitions," *Oikos*, vol. 122, no. 5, pp. 641–648, 2013.

[136] P. Almaraz and A. J. Green, "Catastrophic bifurcation in the dynamics of a threatened bird community triggered by a planetary-scale environmental perturbation," *Biological Conservation*, vol. 291, p. 110466, Mar. 2024.

[137] M. C. Boerlijst, T. Oudman, and A. M. de Roos, "Catastrophic Collapse Can Occur without Early Warning: Examples of Silent Catastrophes in Structured Ecological Models," *PLOS ONE*, vol. 8, no. 4, p. e62033, Apr. 2013.

[138] V. Dakos, "Identifying best-indicator species for abrupt transitions in multispecies communities," *Ecological Indicators*, vol. 94, pp. 494–502, Nov. 2018.

[139] T. Smith, D. Traxl, and N. Boers, "Empirical evidence for recent global shifts in vegetation resilience," *Nature Climate Change*, vol. 12, no. 5, pp. 477–484, May 2022.

[140] J. L. Deichmann, O. Acevedo-Charry, L. Barclay, Z. Burivalova, M. Campos-Cerqueira, F. d'Horta, E. T. Game, B. L. Gottesman, P. J. Hart, A. K. Kalan, S. Linke, L. D. Nascimento, B. Pijanowski, E. Staaterman, and T. Mitchell Aide, "It's time to listen: There is much to be learned from the sounds of tropical ecosystems," *Biotropica*, vol. 50, no. 5, pp. 713–718, 2018.

[141] N. Yoh, W. Mbamy, B. L. Gottesman, G. Z. L. Froese, T. Satchivi, M. Obiang Ebanega, L. Carlson, S. E. Koto, M. Özdoğan, D. J. I. Seaman, V. Maicher, H. Malinowski, J. Poulsen, A. Ebang Mbélé, and Z. Buřivalová, "Impacts of logging, hunting, and conservation on vocalizing biodiversity in Gabon," *Biological Conservation*, vol. 296, p. 110726, Aug. 2024.

[142] D. Winiarska, P. Szymański, and T. S. Osiejuk, "Detection ranges of forest bird vocalisations: Guidelines for passive acoustic monitoring," *Scientific Reports*, vol. 14, no. 1, p. 894, Jan. 2024.

[143] D. Mann, A. Anderson, A. Donner, M. Hall, S. Kahl, and H. Klinck, "Continental-scale behavioral response of birds to a total solar eclipse," *Scientific Reports*, vol. 15, no. 1, p. 10113, Apr. 2025.

[144] O. C. Metcalf, J. Barlow, C. Devenish, S. Marsden, E. Berenguer, and A. C. Lees, "Acoustic indices perform better when applied at ecologically meaningful time and frequency scales," *Methods in Ecology and Evolution*, vol. 12, no. 3, pp. 421–431, 2021.

[145] A. Digby, M. Towsey, B. D. Bell, and P. D. Teal, "A practical comparison of manual and autonomous methods for acoustic monitoring," *Methods in Ecology and Evolution*, vol. 4, no. 7, pp. 675–683, 2013.

[146] D. Stowell, "Computational bioacoustics with deep learning: A review and roadmap," *PeerJ*, vol. 10, p. e13152, Mar. 2022.

[147] N. P. Le Penru, B. E. Heath, J. Dunning, L. Picinali, R. M. Ewers, and S. S. Sethi, "Towards using virtual acoustics for evaluating spatial ecoacoustic monitoring technologies," *Methods in Ecology and Evolution*, vol. 16, no. 1, pp. 108–125, 2025.

[148] B. Holgate, R. Maggini, and S. Fuller, "Mapping ecoacoustic hot spots and moments of biodiversity to inform conservation and urban planning," *Ecological Indicators*, vol. 126, p. 107627, Jul. 2021.

[149] T. Bradfer-Lawrence, B. Duthie, C. Abrahams, M. Adam, R. J. Barnett, A. Beeston, J. Darby, B. Dell, N. Gardner, A. Gasc, B. Heath, N. Howells, M. Janson, M.-V. Kyoseva, T. Luypaert, O. C. Metcalf, A. E. Nousek-McGregor, F. Poznansky, S. R. P.-J. Ross, S. Sethi, S. Smyth, E. Waddell, and J. S. P. Froidevaux, "The Acoustic Index User's Guide: A practical manual for defining, generating and understanding current and future acoustic indices," *Methods in Ecology and Evolution*, vol. 16, no. 6, May 2024.

[150] J. Sueur, A. Farina, A. Gasc, N. Pieretti, and S. Pavoine, "Acoustic Indices for Biodiversity Assessment and Landscape Investigation," *Acta Acustica united with Acustica*, vol. 100, no. 4, pp. 772–781, Jul. 2014.

[151] R. T. Buxton, M. F. McKenna, M. Clapp, E. Meyer, E. Stabenau, L. M. Angeloni, K. Crooks, and G. Wittemyer, "Efficacy of extracting indices from large-scale acoustic recordings to monitor biodiversity," *Conservation Biology*, vol. 32, no. 5, pp. 1174–1184, 2018.

[152] N. Pieretti, A. Farina, and D. Morri, "A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI)," *Ecological Indicators*, vol. 11, no. 3, pp. 868–873, May 2011.

[153] O. Metcalf, C. Abrahams, B. Ashington, E. Baker, T. Bradfer-Lawrence, E. Browning, J. Carruthers-Jones, J. Darby, J. Dick, A. Eldridge, D. Elliott, B. Heath, P. Howden-Leach, A. Johnston, A. Lees, C. Meyer, U. Ruiz Arana, and S. Smyth, "Good practice guidelines for long-term ecoacoustic monitoring in the UK," The UK Acoustics Network, Report, Feb. 2023.

[154] N. T. Boelman, G. P. Asner, P. J. Hart, and R. E. Martin, "Multi-Trophic Invasion Resistance In Hawaii: Bioacoustics, Field Surveys, and Airborne Remote Sensing," *Ecological Applications*, vol. 17, no. 8, pp. 2137–2144, Dec. 2007.

[155] E. P. Kasten, S. H. Gage, J. Fox, and W. Joo, "The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology," *Ecological Informatics*, vol. 12, pp. 50–67, Nov. 2012.

[156] A. Eldridge, P. Guyot, P. Moscoso, A. Johnston, Y. Eyre-Walker, and M. Peck, "Sounding out ecoacoustic metrics: Avian species richness is predicted by acoustic indices in temperate but not tropical habitats," *Ecological Indicators*, vol. 95, pp. 939–952, Dec. 2018.

[157] T. Bradfer-Lawrence, N. Bunnefeld, N. Gardner, S. G. Willis, and D. H. Dent, "Rapid assessment of avian species richness and abundance using acoustic indices," *Ecological Indicators*, vol. 115, p. 106400, Aug. 2020.

[158] S. L. Mitchell, J. E. Bicknell, D. P. Edwards, N. J. Deere, H. Bernard, Z. G. Davies, and M. J. Struebig, "Spatial replication and habitat context matters for assessments of tropical biodiversity using acoustic indices," *Ecological Indicators*, vol. 119, p. 106717, Dec. 2020.

[159] S. S. Sethi, A. Bick, R. M. Ewers, H. Klinck, V. Ramesh, M.-N. Tuanmu, and D. A. Coomes, "Limits to the accurate and generalizable use of soundscapes to monitor biodiversity," *Nature Ecology & Evolution*, vol. 7, no. 9, pp. 1373–1378, Sep. 2023.

[160] C. M. Wood and M. Z. Peery, "What does 'occupancy' mean in passive acoustic surveys?" *Ibis*, vol. 164, no. 4, pp. 1295–1300, 2022.

[161] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778.

[162] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Apr. 2015.

[163] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 131–135.

[164] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*. PMLR, May 2019, pp. 6105–6114.

[165] M. J. Guerrero, C. L. Bedoya, J. D. López, J. M. Daza, and C. Isaza, "Acoustic animal identification using unsupervised learning," *Methods in Ecology and Evolution*, vol. 14, no. 6, pp. 1500–1514, 2023.

[166] S. E. Newson and G. Crisp, *Bat Distribution and Activity in the Skell Valley Catchment, 2024 Report*, ser. BTO Research Report. Thetford, UK: BTO, 2024, no. 776.

[167] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian diversity monitoring," *Ecological Informatics*, vol. 61, p. 101236, Mar. 2021.

[168] C. M. Wood, K. , Stefan, B. , Stephanie, V. H. , Rachel, and C. and Brown, "Passive acoustic surveys and the BirdNET algorithm reveal detailed spatiotemporal

variation in the vocal activity of two anurans," *Bioacoustics*, vol. 32, no. 5, pp. 532–543, Sep. 2023.

[169] J. Hamer, E. Triantafillou, B. van Merriënboer, S. Kahl, H. Klinck, T. Denton, and V. Dumoulin, "BIRB: A Generalization Benchmark for Information Retrieval in Bioacoustics," Dec. 2023.

[170] C. Pérez-Granados, "BirdNET: Applications, performance, pitfalls and future opportunities," *Ibis*, vol. 165, no. 3, Jul. 2023.

[171] I. Tolkova and H. Klinck, "Source separation with an acoustic vector sensor for terrestrial bioacoustics," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, p. 1123, Aug. 2022.

[172] S. S. Sethi, A. Bick, M.-Y. Chen, R. Crouzeilles, B. V. Hillier, J. Lawson, C.-Y. Lee, S.-H. Liu, C. H. de Freitas Parruco, C. M. Rosten, M. Somveille, M.-N. Tuanmu, and C. Banks-Leite, "Large-scale avian vocalization detection delivers reliable global biodiversity insights," *Proceedings of the National Academy of Sciences*, vol. 121, no. 33, p. e2315933121, Aug. 2024.

[173] B. van Merriënboer, J. Hamer, V. Dumoulin, E. Triantafillou, and T. Denton, "Birds, bats and beyond: Evaluating generalization in bioacoustics models," *Frontiers in Bird Science*, vol. 3, Jul. 2024.

[174] B. Ghani, T. Denton, S. Kahl, and H. Klinck, "Global birdsong embeddings enable superior transfer learning for bioacoustic classification," *Scientific Reports*, vol. 13, no. 1, p. 22876, Dec. 2023.

[175] I. Nolasco, S. Singh, V. Morfi, V. Lostanlen, A. Strandburg-Peshkin, E. Vidaña-Vila, L. Gill, H. Pamuła, H. Whitehead, I. Kiskin, F. H. Jensen, J. Morford, M. G. Emmerson, E. Versace, E. Grout, H. Liu, B. Ghani, and D. Stowell, "Learning to detect an animal sound from five examples," *Ecological Informatics*, vol. 77, p. 102258, Nov. 2023.

[176] B. E. Heath, S. S. Sethi, C. D. L. Orme, R. M. Ewers, and L. Picinali, "How index selection, compression, and recording schedule impact the description of ecological soundscapes," *Ecology and Evolution*, vol. 11, no. 19, pp. 13 206–13 217, 2021.

[177] Y. Y. Loo, M. Y. Lee, S. Shaheed, T. Maul, and D. J. Clink, "Temporal patterns in Malaysian rainforest soundscapes demonstrated using acoustic indices and deep embeddings trained on time-of-day estimationa)," *The Journal of the Acoustical Society of America*, vol. 157, no. 1, pp. 1–16, Jan. 2025.

[178] S. S. Sethi, R. M. Ewers, N. S. Jones, J. Sleutel, A. Shabrani, N. Zulkifli, and L. Picinali, "Soundscapes predict species occurrence in tropical forests," *Oikos*, vol. 2022, no. 3, p. e08525, 2022.

[179] I. Tolkova, "Feature Representations for Conservation Bioacoustics: Review and Discussion," in *IJCAI 2021 Workshop on AI for Social Good*, 2021.

[180] W. De Keersmaecker, S. Lhermitte, O. Honnay, J. Farifteh, B. Somers, and P. Coppin, "How to measure ecosystem stability? An evaluation of the reliability of stability metrics based on remote sensing time series across the major global ecosystems," *Global Change Biology*, vol. 20, no. 7, pp. 2149–2161, 2014.

[181] N. Howells, A. C. Lees, J. Barlow, E. Berenguer, L. C. Rossi, J. Sueur, M. J. Sullivan, Y. Gabriel Ramos, and O. C. Metcalf, "Temporal patterns of Amazonian insect acoustic activity," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 380, no. 1928, p. 20240337, Jun. 2025.

[182] M. Padgham, "Reverberation and frequency attenuation in forests—implications for acoustic communication in animals," *The Journal of the Acoustical Society of America*, vol. 115, no. 1, pp. 402–410, Dec. 2003.

[183] E. Knight, K. Hannah, G. Foley, C. Scott, R. Brigham, and E. Bayne, "Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs," *Avian Conservation and Ecology*, vol. 12, no. 2, Nov. 2017.

[184] C. M. Wood and S. Kahl, "Guidelines for appropriate use of BirdNET scores and other detector outputs," *Journal of Ornithology*, Feb. 2024.

[185] V. Dakos, S. R. Carpenter, W. A. Brock, A. M. Ellison, V. Guttal, A. R. Ives, S. Kéfi, V. Livina, D. A. Seekell, E. H. van Nes, and M. Scheffer, "Methods for

Detecting Early Warnings of Critical Transitions in Time Series Illustrated Using Simulated Ecological Data," *PLOS ONE*, vol. 7, no. 7, p. e41010, Jul. 2012.

[186] T. M. Lenton, V. N. Livina, V. Dakos, E. H. van Nes, and M. Scheffer, "Early warning of climate tipping points from critical slowing down: Comparing methods to improve robustness," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1962, pp. 1185–1204, Mar. 2012.

[187] T. M. Bury, R. I. Sujith, I. Pavithran, M. Scheffer, T. M. Lenton, M. Anand, and C. T. Bauch, "Deep learning for early warning signals of tipping points," *Proceedings of the National Academy of Sciences*, vol. 118, no. 39, Sep. 2021.

[188] F. Dablander and T. M. Bury, "Deep learning for tipping points: Preprocessing matters," *Proceedings of the National Academy of Sciences*, vol. 119, no. 37, p. e2207720119, Sep. 2022.

[189] V. Dakos, E. H. van Nes, P. D'Odorico, and M. Scheffer, "Robustness of variance and autocorrelation as indicators of critical slowing down," *Ecology*, vol. 93, no. 2, pp. 264–271, 2012.

[190] T. Smith, R.-M. Zotta, C. A. Boulton, T. M. Lenton, W. Dorigo, and N. Boers, "Reliability of resilience estimation based on multi-instrument time series," *Earth System Dynamics*, vol. 14, no. 1, pp. 173–183, Feb. 2023.

[191] R. Biggs, S. R. Carpenter, and W. A. Brock, "Turning back from the brink: Detecting an impending regime shift in time to avert it," *Proceedings of the National Academy of Sciences*, vol. 106, no. 3, pp. 826–831, Jan. 2009.

[192] V. Guttal and C. Jayaprakash, "Changing skewness: An early warning signal of regime shifts in ecosystems," *Ecology Letters*, vol. 11, no. 5, pp. 450–460, 2008.

[193] V. Dakos, E. H. van Nes, R. Donangelo, H. Fort, and M. Scheffer, "Spatial correlation as leading indicator of catastrophic shifts," *Theoretical Ecology*, vol. 3, no. 3, pp. 163–174, Aug. 2010.

[194] V. N. Livina and T. M. Lenton, "A modified method for detecting incipient bifurcations in a dynamical system," *Geophysical Research Letters*, vol. 34, no. 3, 2007.

[195] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[196] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd ed., 2026, online manuscript released January 6, 2026. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/

[197] K. Vitense, M. A. Hanson, B. R. Herwig, K. D. Zimmer, and J. Fieberg, "Using hidden Markov models to inform conservation and management strategies in ecosystems exhibiting alternative stable states," *Journal of Applied Ecology*, vol. 58, no. 5, pp. 1069–1078, 2021.

[198] C. F. Clements and A. Ozgul, "Including trait-based early warning signals helps predict population collapse," *Nature Communications*, vol. 7, no. 1, p. 10984, Mar. 2016.

[199] C. F. Clements, J. L. Blanchard, K. L. Nash, M. A. Hindell, and A. Ozgul, "Body size shifts and early warning signals precede the historic collapse of whale stocks," *Nature Ecology & Evolution*, vol. 1, no. 7, pp. 1–6, Jun. 2017.

[200] D. L. Bowling, M. Garcia, J. C. Dunn, R. Ruprecht, A. Stewart, K.-H. Frommolt, and W. T. Fitch, "Body size and vocalization in primates and carnivores," *Scientific Reports*, vol. 7, no. 1, p. 41070, Jan. 2017.

[201] K. Martin, M. A. Tucker, and T. L. Rogers, "Does size matter? Examining the drivers of mammalian vocalizations," *Evolution*, vol. 71, no. 2, pp. 249–260, Feb. 2017.

[202] M. Scheffer, S. Barrett, S. R. Carpenter, C. Folke, A. J. Green, M. Holmgren, T. P. Hughes, S. Kosten, I. A. van de Leemput, D. C. Nepstad, E. H. van Nes, E. T. H. M. Peeters, and B. Walker, "Creating a safe operating space for iconic ecosystems," *Science*, vol. 347, no. 6228, pp. 1317–1319, Mar. 2015.

[203] A. C. Lees and P. Yuda, "The Asian songbird crisis," *Current Biology*, vol. 32, no. 20, pp. R1063–R1064, Oct. 2022.

[204] V. Kather, B. Ghani, and D. Stowell, "Clustering and Novel Class Recognition: Evaluating Bioacoustic Deep Learning Feature Extractors," in *Proceedings of the 11th Convention of the European Acoustics Association Forum Acusticum / EuroNoise 2025*. Málaga, Spain: European Acoustics Association, Dec. 2025, pp. 2153–2160.

[205] B. E. Heath, "The assessment and development of methods in (spatial) sound ecology," Ph.D. dissertation, Imperial College London, London, UK, Oct. 2022.

[206] S. Bathiany, M. Claussen, and K. Fraedrich, "Detecting hotspots of atmosphere–vegetation interaction via slowing down - Part 1: A stochastic approach," *Earth System Dynamics*, vol. 4, no. 1, pp. 63–78, Feb. 2013.

[207] F. J. Bravo Sanchez, N. B. English, M. R. Hossain, and S. T. Moore, "Improved analysis of deep bioacoustic embeddings through dimensionality reduction and interactive visualisation," *Ecological Informatics*, vol. 81, p. 102593, Jul. 2024.

[208] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nature Reviews Methods Primers*, vol. 2, no. 1, pp. 1–21, Dec. 2022.

[209] J. J. Lever, I. A. van de Leemput, E. Weinans, R. Quax, V. Dakos, E. H. van Nes, J. Bascompte, and M. Scheffer, "Foreseeing the future of mutualistic communities beyond collapse," *Ecology Letters*, vol. 23, no. 1, pp. 2–15, 2020.

[210] A. P. Hill, P. Prince, J. L. Snaddon, C. P. Doncaster, and A. Rogers, "AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment," *HardwareX*, vol. 6, Oct. 2019.

[211] S. S. Sethi, M. Kovac, F. Wiesemüller, A. Miriyev, and C. M. Boutry, "Biodegradable sensors are ready to transform autonomous ecological monitoring," *Nature Ecology & Evolution*, vol. 6, no. 9, pp. 1245–1247, Sep. 2022.

[212] D. Teixeira, P. Roe, B. J. van Rensburg, S. Linke, P. G. McDonald, D. Tucker, and S. Fuller, "Effective ecological monitoring using passive acoustic sensors: Recommendations for conservation practitioners," *Conservation Science and Practice*, vol. 6, no. 6, p. e13132, 2024.

[213] R. T. Buxton, P. E. Lendrum, K. R. Crooks, and G. Wittemyer, "Pairing camera traps and acoustic recorders to monitor the ecological impact of human disturbance," *Global Ecology and Conservation*, vol. 16, p. e00493, Oct. 2018.

[214] J. Gottwald, P. Lampe, J. Höchst, N. Friess, J. Maier, L. Leister, B. Neumann, T. Richter, B. Freisleben, and T. Nauss, "BatRack: An open-source multi-sensor device for wildlife research," *Methods in Ecology and Evolution*, vol. 12, no. 10, pp. 1867–1874, 2021.

[215] J. W. Wägele, P. Bodesheim, S. J. Bourlat, J. Denzler, M. Diepenbroek, V. Fonseca, K.-H. Frommolt, M. F. Geiger, B. Gemeinholzer, F. O. Glöckner, T. Haucke, A. Kirse, A. Kölpin, I. Kostadinov, H. S. Kühl, F. Kurth, M. Lasseck, S. Liedke, F. Losch, S. Müller, N. Petrovskaya, K. Piotrowski, B. Radig, C. Scherber, L. Schoppmann, J. Schulz, V. Steinhage, G. F. Tschan, W. Vautz, D. Velotto, M. Weigend, and S. Wildermann, "Towards a multisensor station for automated biodiversity monitoring," *Basic and Applied Ecology*, vol. 59, pp. 105–138, Mar. 2022.

[216] A. Bush, R. Sollmann, A. Wilting, K. Bohmann, B. Cole, H. Balzter, C. Martius, A. Zlinszky, S. Calvignac-Spencer, C. A. Cobbold, T. P. Dawson, B. C. Emerson, S. Ferrier, M. T. P. Gilbert, M. Herold, L. Jones, F. H. Leendertz, L. Matthews, J. D. A. Millington, J. R. Olson, O. Ovaskainen, D. Raffaelli, R. Reeve, M.-O. Rödel, T. W. Rodgers, S. Snape, I. Visseren-Hamakers, A. P. Vogler, P. C. L. White, M. J. Wooster, and D. W. Yu, "Connecting Earth observation to high-throughput biodiversity data," *Nature Ecology & Evolution*, vol. 1, no. 7, pp. 1–9, Jun. 2017.

[217] R. M. Ewers, J. Cook, O. Z. Daniel, C. D. L. Orme, V. Groner, J. Joshi, A. Rallings, T. Rallings, and P. Amarasekare, "New insights to be gained from a Virtual Ecosystem," *Ecological Modelling*, vol. 498, p. 110866, Dec. 2024.

[218] S. L. Pimm, "The complexity and stability of ecosystems," *Nature*, vol. 307, no. 5949, pp. 321–326, Jan. 1984.

[219] B. Walker, C. S. Holling, S. Carpenter, and A. Kinzig, "Resilience, Adaptability and Transformability in Social–ecological Systems," *Ecology and Society*, vol. 9, no. 2, Sep. 2004.

[220] A. Hastings, "Transients: The key to long-term ecological understanding?" *Trends in Ecology & Evolution*, vol. 19, no. 1, pp. 39–45, Jan. 2004.

[221] R. M. May, D. M. G. Wishart, J. Bray, R. L. Smith, M. V. Berry, I. C. Percival, and N. O. Weiss, "Chaos and the dynamics of biological populations," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 413, no. 1844, pp. 27–44, Jan. 1987.

[222] A. Hastings, C. L. Hom, S. Ellner, P. Turchin, and H. C. J. Godfray, "Chaos in Ecology: Is Mother Nature a Strange Attractor?*," *Annual Review of Ecology, Evolution, and Systematics*, vol. 24, no. Volume 24, 1993, pp. 1–33, Nov. 1993.

[223] T. H. Oliver, M. S. Heard, N. J. B. Isaac, D. B. Roy, D. Procter, F. Eigenbrod, R. Freckleton, A. Hector, C. D. L. Orme, O. L. Petchey, V. Proença, D. Raffaelli, K. B. Suttle, G. M. Mace, B. Martín-López, B. A. Woodcock, and J. M. Bullock, "Biodiversity and Resilience of Ecosystem Functions," *Trends in Ecology & Evolution*, vol. 30, no. 11, pp. 673–684, Nov. 2015.

[224] T. C. Collier, A. N. G. Kirschel, and C. E. Taylor, "Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network," *The Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. 182–189, Jul. 2010.

[225] D. J. Mennill and S. L. Vehrencamp, "Context-Dependent Functions of Avian Duets Revealed by Microphone-Array Recordings and Multispeaker Playback," *Current Biology*, vol. 18, no. 17, pp. 1314–1319, Sep. 2008.

[226] K. A. Hobson, R. S. Rempel, H. Greenwood, B. Turnbull, and S. L. Van Wilgenburg, "Acoustic Surveys of Birds Using Electronic Recordings: New Potential from an Omnidirectional Microphone System," *Wildlife Society Bulletin (1973-2006)*, vol. 30, no. 3, pp. 709–720, 2002.

[227] D. Payo, L. González, P. Kogan, and M. Eguia, "Sentinel: Versatile real-time acoustic autonomous monitoring system for studying natural soundscapes," in *Euronoise 2021*, Madeira, Portugal, Oct. 2021.

[228] A. Celis-Murillo, J. L. Deppe, and M. F. Allen, "Using soundscape recordings to estimate bird species abundance, richness, and composition," *Journal of Field Ornithology*, vol. 80, no. 1, pp. 64–78, 2009.

[229] M. Wijers, A. Loveridge, D. W. Macdonald, and A. Markham, "CARACAL: A versatile passive acoustic monitoring tool for wildlife research and conservation," *Bioacoustics*, vol. 30, no. 1, pp. 41–57, Nov. 2019.

[230] B. Cretois, C. M. Rosten, and S. S. Sethi, "Voice activity detection in eco-acoustic data enables privacy protection and is a proxy for human disturbance," *Methods in Ecology and Evolution*, vol. 13, no. 12, pp. 2865–2874, 2022.

[231] S. Kaneko and H. Gamper, "Large-scale simulation of bird localization systems in forests with distributed microphone arrays: JASA Express Letters: Vol 2, No 10," *JASA Express Letters*, vol. 2, no. 10, pp. 101 201–1 – 101 201–8, Oct. 2022.

[232] C. Oreinos and J. M. Buchholz, "Evaluation of Loudspeaker-Based Virtual Sound Environments for Testing Directional Hearing Aids," *Journal of the American Academy of Audiology*, vol. 27, no. 07, pp. 541–556, Jul. 2016.

[233] N. Mansour, M. Marschall, A. Westermann, T. May, and T. Dau, "Speech Intelligibility in a Realistic Virtual Sound Environment: 23rd International Congress on Acoustics," in *Proceedings of the 23rd International Congress on Acoustics*. Aachen, Germany: Deutsche Gesellschaft für Akustik, Sep. 2019, pp. 7658–7665.

[234] L. S. R. Simon, H. Wüthrich, and N. Dillier, "Investigating Higher Order Spatial Impulse Response Rendering for the evaluation of hearing devices," in *DAGA 2021*, Vienna, Austria, 2021.

[235] S. Favrot and J. M. Buchholz, "LoRA: A Loudspeaker-Based Room Auralization System," *Acta Acustica united with Acustica*, vol. 96, no. 2, pp. 364–375, Mar. 2010.

[236] C. T. J. Hui, E. Au, S. Xiao, Y. Hioka, H. Masuda, and C. I. Watson, "Differences in speech intelligibility in noise between native and non-native listeners under ambisonics-based sound reproduction system," *Applied Acoustics*, vol. 184, p. 108368, Dec. 2021.

[237] C. Müller, *Spherical Harmonics*. Springer, 2006, vol. 17.

[238] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université Paris VI, Paris, France, Jul. 2001.

[239] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of Spherical Microphone," in *Audio Engineering Society Convention 120*, Paris, France, May 2006.

[240] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, ser. Springer Topics in Signal Processing. Cham: Springer International Publishing, 2019, vol. 19.

[241] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc*, vol. 21, no. 1, pp. 2–10, 1973.

[242] ——, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," in *Audio Engineering Society Convention 50*, Mar. 1975.

[243] M. Frank, "How to make Ambisonics sound good," in *Forum Acusticum 2014*, Krakow, Poland, 2014.

[244] N. Epain and J. Daniel, "Improving Spherical Microphone Arrays," in *Audio Engineering Society Convention 124*, May 2008.

[245] M. A. Gerzon, "General Metatheory of Auditory Localisation," in *Audio Engineering Society Convention 92*, Mar. 1992.

[246] E. Quirós-Guerrero, M. J. Janeiro, M. Lopez-Morales, W. Cresswell, and C. N. Templeton, "Riverside wren pairs jointly defend their territories against simulated intruders," *Ethology*, vol. 123, no. 12, pp. 949–956, 2017.

[247] D. Monacchi, "Fragments of Extinction: Acoustic Biodiversity of Primary Rainforest Ecosystems," *Leonardo Music Journal*, vol. 23, pp. 23–25, Dec. 2013.

[248] Zylia Sp. z o.o., "ZYLIA ZM-1 microphone," Poznań, Poland. [Online]. Available: https://www.zylia.co/zylia-zm-1-microphone.html

[249] ——, "ZYLIA Studio," Poznań, Poland, Jan. 2022. [Online]. Available: https://www.zylia.co/zylia-studio.html

[250] ——, "ZYLIA Ambisonics Converter," Poznań, Poland, Jan. 2022. [Online]. Available: https://www.zylia.co/zylia-ambisonics-converter-download.html

[251] Institute for Computer Music and Sound Technology, "Ambisonics Externals for MaxMSP," Zurich University of the Arts, Zurich, Switzerland, Nov. 2021. [Online]. Available: https://www.zhdk.ch/forschung/icst/software-downloads-5379/downloads-ambisonics-externals-for-maxmsp-5381

[252] Seeed Studio, "ReSpeaker 6-Mic Circular Array kit for Raspberry Pi." [Online]. Available: https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/

[253] R. B. Machado, L. Aguiar, and G. Jones, "Do acoustic indices reflect the characteristics of bird communities in the savannas of Central Brazil?" *Landscape and Urban Planning*, vol. 162, pp. 36–43, Jun. 2017.

[254] T. Bradfer-Lawrence, N. Gardner, L. Bunnefeld, N. Bunnefeld, S. G. Willis, and D. H. Dent, "Guidelines for the use of acoustic indices in environmental research," *Methods in Ecology and Evolution*, vol. 10, no. 10, pp. 1796–1807, 2019.

[255] J. Sueur, T. Aubin, and C. Simonis, "Seewave, a Free Modular Tool for Sound Analysis and Synthesis," *Bioacoustics*, vol. 18, no. 2, pp. 213–226, Jan. 2008.

[256] B. C. Pijanowski and L. J. Villanueva-Rivera, "Soundecology: Soundscape Ecology," Mar. 2018. [Online]. Available: https://cran.r-project.org/web/packages/soundecology/

[257] The MathWorks, Inc., "MATLAB Deep Learning Toolbox," 2022. [Online]. Available: https://uk.mathworks.com/help/deeplearning/

[258] D. Giavarina, "Understanding Bland Altman analysis," *Biochemia Medica*, vol. 25, no. 2, pp. 141–151, Jun. 2015.

[259] M. Araya-Salas, G. Smith-Vidaurre, and M. Webster, "Assessing the effect of sound file compression and background noise on measures of acoustic signal structure," *Bioacoustics*, vol. 28, no. 1, pp. 57–73, Jan. 2019.

[260] N. Krabbe, "XC25554 Common Blackbird (Turdus merula)," Jun. 1997. [Online]. Available: https://xeno-canto.org/25554

[261] P. Ålberg, "XC28208 Eurasian Blue Tit (Cyanistes caeruleus)," May 2006. [Online]. Available: https://xeno-canto.org/28208

[262] ——, "XC27554 European Robin (Erithacus rubecula)," Jun. 2008. [Online]. Available: https://xeno-canto.org/27554

[263] S. Bot, "XC35048 European Greenfinch (Chloris chloris)," Apr. 2009. [Online]. Available: https://xeno-canto.org/35048

[264] J. Matusiak, "XC44727 House Sparrow (Passer domesticus)," Feb. 2010. [Online]. Available: https://xeno-canto.org/44727

[265] J. Poelstra, "XC76918 Common Wood Pigeon (Columba palumbus)," Apr. 2011. [Online]. Available: https://xeno-canto.org/76918

[266] L. Picinali, "Techniques for the extraction of the impulse response of a linear and time-invariant system." in *Proceedings of the DMRN Doctoral Research Conference, Goldsmiths College, University of London, UK, 22nd-23rd July 2006*. London, UK: DMRN Doctoral Researchers Conference, Goldsmith College, University of London, Jul. 2006.

[267] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio Set: An ontology and human-labeled dataset for audio events," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 776–780.

[268] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc*, vol. 45, no. 6, pp. 456–466, 1997.

[269] T.-H. Lin and Y. Tsao, "Source separation in ecoacoustics: A roadmap towards versatile soundscape information retrieval," *Remote Sensing in Ecology and Conservation*, vol. 6, no. 3, pp. 236–247, 2020.

[270] M. R. Izadi, R. Stevenson, and L. N. Kloepper, "Separation of overlapping sources in bioacoustic mixtures," *The Journal of the Acoustical Society of America*, vol. 147, no. 3, pp. 1688–1696, Mar. 2020.

[271] P. C. Bermant, "BioCPPNet: Automatic bioacoustic source separation with deep neural networks," *Scientific Reports*, vol. 11, no. 1, p. 23502, Dec. 2021.

[272] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[273] D. L. Jones and R. Ratnam, "Blind location and separation of callers in a natural chorus using a microphone array," *The Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 895–910, Aug. 2009.

[274] I. McCowan, "Microphone arrays and beamforming," in *Multimodal Signal Processing: Human Interactions in Meetings*, S. Renals, H. Bourlard, J. Carletta, and A. Popescu-Belis, Eds. Cambridge: Cambridge University Press, 2012, pp. 28–39.

[275] S. Cai, T. Collier, L. Girod, J. Y. Lee, R. E. Hudson, K. Yao, C. E. Taylor, M. Bao, and Z. Wang, "New wireless acoustic array node for localization, beamforming and source separation for bio-complexity bird data collection and study," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, Jul. 2013, pp. 210–214.

[276] R. Kojima, O. Sugiyama, K. Hoshiba, K. Nakadai, R. Suzuki, and C. E. Taylor, "Bird Song Scene Analysis Using a Spatial-Cue-Based Probabilistic Model," *Journal of Robotics and Mechatronics*, vol. 29, no. 1, pp. 236–246, Feb. 2017.

[277] LABmaker, "SongBeam." [Online]. Available: https://www.labmaker.org/products/songbeam

[278] M. J. Anderson, T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J. Sanders, H. V. Cornell, L. S. Comita, K. F. Davies, S. P. Harrison, N. J. B. Kraft, J. C. Stegen, and N. G. Swenson, "Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist," *Ecology Letters*, vol. 14, no. 1, pp. 19–28, Jan. 2011.

[279] E. Sirjani, P. J. Cragg, and M. K. Dymond, "Glass transition temperatures, melting temperatures, water contact angles and dimensional precision of simple fused deposition model 3D prints and 3D printed channels constructed from a range of

commercially available filaments," *Chemical Data Collections*, vol. 22, p. 100244, Aug. 2019.

[280] The MathWorks, Inc., "Element and Array Radiation and Response Patterns - MATLAB & Simulink." [Online]. Available: https://uk.mathworks.com/help/phased/ug/element-and-array-radiation-patterns-and-responses.html

[281] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA: John Wiley & Sons, Inc., Feb. 2016.

[282] F. Grondin, D. Létourneau, C. Godin, J.-S. Lauzon, J. Vincent, S. Michaud, S. Faucher, and F. Michaud, "ODAS: Open embeddeD Audition System," *Frontiers in Robotics and AI*, vol. 9, 2022.

[283] C. M. Wood, S. Kahl, P. Chaon, M. Z. Peery, and H. Klinck, "Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys," *Methods in Ecology and Evolution*, vol. 12, no. 5, pp. 885–896, 2021.

[284] J. L. Rochat and D. Reiter, "Highway Traffic Noise," *Acoustics Today*, vol. 12, no. 4, pp. 38–47, 2016.

[285] G. Porter, "Lithobates clamitans," 2023. [Online]. Available: https://animaldiversity.org/accounts/Lithobates_clamitans/

[286] A. Pitt, J. Tavano, R. Baldwin, and B. Stegenga, "Movement ecology and habitat use of three sympatric anuran species," *Herpetological Conservation and Biology*, vol. 12, pp. 212–224, Apr. 2017.

[287] C. J. Marsh, E. C. Turner, B. W. Blonder, B. Bongalov, S. Both, R. S. Cruz, D. M. O. Elias, D. Hemprich-Bennett, P. Jotan, V. Kemp, U. H. Kritzler, S. Milne, D. T. Milodowski, S. L. Mitchell, M. M. Pillco, M. H. Nunes, T. Riutta, S. J. B. Robinson, E. M. Slade, H. Bernard, D. F. R. P. Burslem, A. Y. C. Chung, E. L. Clare, D. A. Coomes, Z. G. Davies, D. P. Edwards, D. Johnson, P. Kratina, Y. Malhi, N. Majalap, R. Nilus, N. J. Ostle, S. J. Rossiter, M. J. Struebig, J. A. Tobias, M. Williams, R. M. Ewers, O. T. Lewis, G. Reynolds, Y. A. Teh, and A. Hec-

tor, "Tropical forest clearance impacts biodiversity and function, whereas logging changes structure," *Science*, vol. 387, no. 6730, pp. 171–175, Jan. 2025.

[288] S. F. B. Ferraz, K. M. P. M. B. Ferraz, C. C. Cassiano, P. H. S. Brancalion, D. T. A. da Luz, T. N. Azevedo, L. R. Tambosi, and J. P. Metzger, "How good are tropical forest patches for ecosystem services provisioning?" *Landscape Ecology*, vol. 29, no. 2, pp. 187–200, Feb. 2014.

[289] T. A. Gardner, J. Barlow, R. Chazdon, R. M. Ewers, C. A. Harvey, C. A. Peres, and N. S. Sodhi, "Prospects for tropical forest biodiversity in a human-modified world," *Ecology Letters*, vol. 12, no. 6, pp. 561–582, Jun. 2009.

[290] R. A. Houghton, B. Byers, and A. A. Nassikas, "A role for tropical forests in stabilizing atmospheric CO2," *Nature Climate Change*, vol. 5, no. 12, pp. 1022–1023, Dec. 2015.

[291] Y. Pan, R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, A. D. McGuire, S. Piao, A. Rautiainen, S. Sitch, and D. Hayes, "A Large and Persistent Carbon Sink in the World's Forests," *Science*, vol. 333, no. 6045, pp. 988–993, Aug. 2011.

[292] C. Y. Shimamoto, A. A. Padial, C. M. da Rosa, and M. C. M. Marques, "Restoration of ecosystem services in tropical forests: A global meta-analysis," *PLOS ONE*, vol. 13, no. 12, p. e0208523, Dec. 2018.

[293] F. Achard, H. D. Eva, H.-J. Stibig, P. Mayaux, J. Gallego, T. Richards, and J.-P. Malingreau, "Determination of Deforestation Rates of the World's Humid Tropical Forests," *Science*, vol. 297, no. 5583, pp. 999–1002, Aug. 2002.

[294] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend, "High-Resolution Global Maps of 21st-Century Forest Cover Change," *Science*, vol. 342, no. 6160, pp. 850–853, Nov. 2013.

[295] L. M. Curran, S. N. Trigg, A. K. McDonald, D. Astiani, Y. M. Hardiono, P. Siregar, I. Caniago, and E. Kasischke, "Lowland Forest Loss in Protected Areas of Indonesian Borneo," *Science*, vol. 303, no. 5660, pp. 1000–1003, Feb. 2004.

[296] D. L. A. Gaveau, D. Sheil, Husnayaen, M. A. Salim, S. Arjasakusuma, M. Ancrenaz, P. Pacheco, and E. Meijaard, "Rapid conversions and avoided deforestation: Examining four decades of industrial plantation expansion in Borneo," *Scientific Reports*, vol. 6, no. 1, p. 32017, Sep. 2016.

[297] A. Langner and F. Siegert, "Spatiotemporal fire occurrence in Borneo over a period of 10 years," *Global Change Biology*, vol. 15, no. 1, pp. 48–62, 2009.

[298] R. M. May, "Will a Large Complex System be Stable?" *Nature*, vol. 238, no. 5364, pp. 413–414, Aug. 1972.

[299] S. Kéfi, V. Domínguez-García, I. Donohue, C. Fontaine, E. Thébault, and V. Dakos, "Advancing our understanding of ecological stability," *Ecology Letters*, vol. 22, no. 9, pp. 1349–1356, 2019.

[300] C. Coux, R. Rader, I. Bartomeus, and J. M. Tylianakis, "Linking species functional roles to their network roles," *Ecology Letters*, vol. 19, no. 7, pp. 762–770, 2016.

[301] T. Elmqvist, C. Folke, M. Nyström, G. Peterson, J. Bengtsson, B. Walker, and J. Norberg, "Response diversity, ecosystem change, and resilience," *Frontiers in Ecology and the Environment*, vol. 1, no. 9, pp. 488–494, 2003.

[302] D. Mouillot, N. A. J. Graham, S. Villéger, N. W. H. Mason, and D. R. Bellwood, "A functional approach reveals community responses to disturbances," *Trends in Ecology & Evolution*, vol. 28, no. 3, pp. 167–177, Mar. 2013.

[303] V. Dakos, B. Matthews, A. P. Hendry, J. Levine, N. Loeuille, J. Norberg, P. Nosil, M. Scheffer, and L. De Meester, "Ecosystem tipping points in an evolving world," *Nature Ecology & Evolution*, vol. 3, no. 3, pp. 355–362, Mar. 2019.

[304] B. M. Flores and M. Holmgren, "Why forest fails to recover after repeated wildfires in Amazonian floodplains? Experimental evidence on tree recruitment limitation," *Journal of Ecology*, vol. 109, no. 10, pp. 3473–3486, 2021.

[305] B. M. Flores, M. Holmgren, C. Xu, E. H. van Nes, C. C. Jakovac, R. C. G. Mesquita, and M. Scheffer, "Floodplains as an Achilles' heel of Amazonian forest resilience," *Proceedings of the National Academy of Sciences*, vol. 114, no. 17, pp. 4442–4446, Apr. 2017.

[306] C. A. Boulton, T. M. Lenton, and N. Boers, "Reply to: Little evidence that Amazonian rainforests are approaching a tipping point," *Nature Climate Change*, pp. 1–3, Nov. 2023.

[307] M. Friedl and D. Sulla-Menashe, "MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V061 | NASA Earthdata," Jun. 2025. [Online]. Available: https://www.earthdata.nasa.gov/data/catalog/lpcloud-mcd12c1-061

[308] L. Moesinger, W. Dorigo, R. de Jeu, R. van der Schalie, T. Scanlon, I. Teubner, and M. Forkel, "The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA)," *Earth System Science Data*, vol. 12, no. 1, pp. 177–196, Jan. 2020.

[309] Y. Y. Liu, R. A. M. de Jeu, M. F. McCabe, J. P. Evans, and A. I. J. M. van Dijk, "Global long-term passive microwave satellite-based retrievals of vegetation optical depth," *Geophysical Research Letters*, vol. 38, no. 18, 2011.

[310] Y. Y. Liu, A. I. J. M. van Dijk, R. A. M. de Jeu, J. G. Canadell, M. F. McCabe, J. P. Evans, and G. Wang, "Recent reversal in loss of global terrestrial biomass," *Nature Climate Change*, vol. 5, no. 5, pp. 470–474, May 2015.

[311] M. J. E. van Marle, G. R. van der Werf, R. a. M. de Jeu, and Y. Y. Liu, "Annual South American forest loss estimates based on passive microwave remote sensing (1990–2010)," *Biogeosciences*, vol. 13, no. 2, pp. 609–624, Feb. 2016.

[312] F. Frappart, J.-P. Wigneron, X. Li, X. Liu, A. Al-Yaari, L. Fan, M. Wang, C. Moisy, E. Le Masson, Z. Aoulad Lafkih, C. Vallé, B. Ygorra, and N. Baghdadi, "Global Monitoring of the Vegetation Dynamics from the Vegetation Optical Depth (VOD): A Review," *Remote Sensing*, vol. 12, no. 18, p. 2915, Jan. 2020.

[313] A. G. Konings, K. Rao, and S. C. Steele-Dunne, "Macro to micro: Microwave remote sensing of plant water content for physiology and ecology," *New Phytologist*, vol. 223, no. 3, pp. 1166–1172, 2019.

[314] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, and J. Michaelsen, "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes," *Scientific Data*, vol. 2, no. 1, p. 150066, Dec. 2015.

[315] Copernicus Climate Change Service, "ERA5-Land monthly averaged data from 1950 to present," 2019. [Online]. Available: https://cds.climate.copernicus.eu/doi/10.24381/cds.68d2bb30

[316] H. Mu, X. Li, Y. Wen, J. Huang, P. Du, W. Su, S. Miao, and M. Geng, "A global record of annual terrestrial Human Footprint dataset from 2000 to 2018," *Scientific Data*, vol. 9, no. 1, p. 176, Apr. 2022.

[317] IUCN, "The IUCN Red List of Threatened Species," 2025. [Online]. Available: https://www.iucnredlist.org/en

[318] R. J. Hijmans, M. Barbosa, R. Bivand, A. Brown, M. Chirico, E. Cordano, K. Dyba, E. Pebesma, B. Rowlingson, and M. D. Sumner, "Terra: Spatial Data Analysis," Sep. 2025. [Online]. Available: https://cran.r-project.org/web/packages/terra/index.html

[319] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria, 2021.

[320] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[321] R. S. Tsay and G. C. Tiao, "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models," *Journal of the American Statistical Association*, vol. 79, no. 385, pp. 84–96, Mar. 1984.

[322] C. A. Boulton, P. Good, and T. M. Lenton, "Early warning signals of simulated Amazon rainforest dieback," *Theoretical Ecology*, vol. 6, no. 3, pp. 373–384, Aug. 2013.

[323] J. Lindeløv, "Mcp: An R Package for Regression With Multiple Change Points," Jan. 2020.

[324] N. Boers, "Early-warning signals for Dansgaard-Oeschger events in a high-resolution ice core record," *Nature Communications*, vol. 9, no. 1, p. 2556, Jul. 2018.

[325] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, "Testing for nonlinearity in time series: The method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1, pp. 77–94, Sep. 1992.

[326] T. A. Gardner, J. Ferreira, J. Barlow, A. C. Lees, L. Parry, I. C. G. Vieira, E. Berenguer, R. Abramovay, A. Aleixo, C. Andretti, L. E. O. C. Aragão, I. Araújo, W. S. de Ávila, R. D. Bardgett, M. Batistella, R. A. Begotti, T. Beldini, D. E. de Blas, R. F. Braga, D. d. L. Braga, J. G. de Brito, P. B. de Camargo, F. Campos dos Santos, V. C. de Oliveira, A. C. N. Cordeiro, T. M. Cardoso, D. R. de Carvalho, S. A. Castelani, J. C. M. Chaul, C. E. Cerri, F. d. A. Costa, C. D. F. da Costa, E. Coudel, A. C. Coutinho, D. Cunha, Á. D'Antona, J. Dezincourt, K. Dias-Silva, M. Durigan, J. C. D. M. Esquerdo, J. Feres, S. F. d. B. Ferraz, A. E. d. M. Ferreira, A. C. Fiorini, L. V. F. da Silva, F. S. Frazão, R. Garrett, A. d. S. Gomes, K. d. S. Gonçalves, J. B. Guerrero, N. Hamada, R. M. Hughes, D. C. Igliori, E. d. C. Jesus, L. Juen, M. Junior, J. M. B. d. O. Junior, R. C. d. O. Junior, C. S. Junior, P. Kaufmann, V. Korasaki, C. G. Leal, R. Leitão, N. Lima, M. d. F. L. Almeida, R. Lourival, J. Louzada, R. M. Nally, S. Marchand, M. M. Maués, F. M. S. Moreira, C. Morsello, N. Moura, J. Nessimian, S. Nunes, V. H. F. Oliveira, R. Pardini, H. C. Pereira, P. S. Pompeu, C. R. Ribas, F. Rossetti, F. A. Schmidt, R. da Silva, R. C. V. M. da Silva, T. F. M. R. da Silva, J. Silveira, J. V. Siqueira, T. S. de Carvalho, R. R. C. Solar, N. S. H. Tancredi, J. R. Thomson, P. C. Torres, F. Z. Vaz-de-Mello, R. C. S. Veiga, A. Venturieri, C. Viana, D. Weinhold, R. Zanetti, and J. Zuanon, "A social and ecological assessment of tropical land uses at multiple scales: The Sustainable Amazon Network," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1619, p. 20120166, Jun. 2013.

[327] T. M. Bury, "Ewstools: A Python package for early warning signals of bifurcations in time series data," *Journal of Open Source Software*, vol. 8, no. 82, p. 5038, Feb. 2023.

[328] A. Hastings, "Timescales, dynamics, and ecological understanding," *Ecology*, vol. 91, no. 12, pp. 3471–3480, 2010.

[329] S. A. Levin, "The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture," *Ecology*, vol. 73, no. 6, pp. 1943–1967, 1992.

[330] C. P. O. Reyer, N. Brouwers, A. Rammig, B. W. Brook, J. Epila, R. F. Grant, M. Holmgren, F. Langerwisch, S. Leuzinger, W. Lucht, B. Medlyn, M. Pfeifer, J. Steinkamp, M. C. Vanderwel, H. Verbeeck, and D. M. Villela, "Forest resilience and tipping points at different spatio-temporal scales: Approaches and challenges," *Journal of Ecology*, vol. 103, no. 1, pp. 5–15, 2015.

[331] J. Aguirre-Gutiérrez, N. Stevens, and E. Berenguer, "Valuing the functionality of tropical ecosystems beyond carbon," *Trends in Ecology & Evolution*, vol. 0, no. 0, Oct. 2023.

[332] C. D. Allen, A. K. Macalady, H. Chenchouni, D. Bachelet, N. McDowell, M. Vennetier, T. Kitzberger, A. Rigling, D. D. Breshears, E. H. T. Hogg, P. Gonzalez, R. Fensham, Z. Zhang, J. Castro, N. Demidova, J.-H. Lim, G. Allard, S. W. Running, A. Semerci, and N. Cobb, "A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests," *Forest Ecology and Management*, vol. 259, no. 4, pp. 660–684, Feb. 2010.

[333] T. T. Kozlowski, "Responses of woody plants to flooding and salinity," *Tree Physiology*, vol. 17, no. 7, p. 490, Jul. 1997.

[334] J. Kreuzwieser and H. Rennenberg, "Molecular and physiological responses of trees to waterlogging stress," *Plant, Cell & Environment*, vol. 37, no. 10, pp. 2245–2259, 2014.

[335] V. R. Milici, D. Dalui, J. G. Mickley, and R. Bagchi, "Responses of plant–pathogen interactions to precipitation: Implications for tropical tree richness in a changing world," *Journal of Ecology*, vol. 108, no. 5, pp. 1800–1809, 2020.

[336] R. J. Morris, "Anthropogenic impacts on tropical forest biodiversity: A network structure and ecosystem functioning perspective," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, no. 1558, pp. 3709–3718, Nov. 2010.

[337] E. L. Neuschulz, T. Mueller, M. Schleuning, and K. Böhning-Gaese, "Pollination and seed dispersal are the most threatened processes of plant regeneration," *Scientific Reports*, vol. 6, no. 1, p. 29839, Jul. 2016.

[338] M. Loreau and C. de Mazancourt, "Biodiversity and ecosystem stability: A synthesis of underlying mechanisms," *Ecology Letters*, vol. 16, no. s1, pp. 106–115, 2013.

[339] T. H. Oliver, N. J. B. Isaac, T. A. August, B. A. Woodcock, D. B. Roy, and J. M. Bullock, "Declining resilience of ecosystem functions under biodiversity loss," *Nature Communications*, vol. 6, no. 1, p. 10122, Dec. 2015.

[340] S. R. Carpenter and W. A. Brock, "Early warnings of regime shifts in spatial dynamics using the discrete Fourier transform," *Ecosphere*, vol. 1, no. 5, p. art10, 2010.

[341] C. T. Perretti and S. B. Munch, "Regime shift indicators fail under noise levels commonly observed in ecological systems," *Ecological Applications*, vol. 22, no. 6, pp. 1772–1779, 2012.

[342] S. Sloan, M. J. Campbell, M. Alamgir, A. M. Lechner, J. Engert, and W. F. Laurance, "Trans-national conservation and infrastructure development in the Heart of Borneo," *PLOS ONE*, vol. 14, no. 9, p. e0221947, Sep. 2019.

[343] J. Varghese and S. S. Crawford, "A cultural framework for Indigenous, Local, and Science knowledge systems in ecology and natural resource management," *Ecological Monographs*, vol. 91, no. 1, p. e01431, 2021.

[344] J. Gazing Wolf, D. D. Ignace, D. M. David-Chavez, L. L. Jennings, D. Smiles, P. Blanchard, E. Simmons, D. Doan-Crider, R. P. S.-S. Kills, M. Montgomery, M. K. Nelson, L. Black Elk, L. Black Elk, G. Bridge, A. M. Chischilly, K. Deer, K. DeerinWater, T. Ecoffey, J. Vergun, D. Wildcat, and J. Rattling Leaf, "Centering Indigenous Knowledges in ecology and beyond," *Frontiers in Ecology and the Environment*, vol. 22, no. 7, p. e2776, 2024.

[345] K. Lisa Yang Center for Conservation Bioacoustics, "BEAT: A bioacoustics learning community for Indonesia and Malaysia," Ithaca, NY, USA. [Online]. Available: https://www.birds.cornell.edu/ccb/beat-a-bioacoustics-learning-community-for-indonesia-and-malaysia/

[346] ——, "Elephant Listening Project," Ithaca, NY, USA. [Online]. Available: https://www.birds.cornell.edu/ccb/elephant-listening-project/

[347] C. A. Boulton, J. E. Buxton, and T. M. Lenton, "Early opportunity signals of a tipping point in the UK's second-hand electric vehicle market," *Earth System Dynamics*, vol. 16, no. 2, pp. 411–421, Mar. 2025.

[348] E. L. Lawrance, R. Thompson, J. Newberry Le Vay, L. Page, and N. Jennings, "The Impact of Climate Change on Mental Health and Emotional Wellbeing: A Narrative Review of Current Evidence, and its Implications," *International Review of Psychiatry*, vol. 34, no. 5, pp. 443–498, Jul. 2022.

[349] A. J. Smalley, M. P. White, R. Ripley, T. X. Atack, E. Lomas, M. Sharples, P. A. Coates, N. Groom, A. Grand, A. Heneberry, L. E. Fleming, and M. H. Depledge, "Forest 404: Using a BBC drama series to explore the impact of nature's changing soundscapes on human wellbeing and behavior," *Global Environmental Change*, vol. 74, p. 102497, May 2022.

[350] C. M. Wood, S. Kahl, A. Rahaman, and H. Klinck, "The machine learning–powered BirdNET App reduces barriers to global bird research by enabling citizen science participation," *PLOS Biology*, vol. 20, no. 6, p. e3001670, Jun. 2022.

[351] T. M. Bury, C. T. Bauch, and M. Anand, "Detecting and distinguishing tipping points using spectral early warning signals," *Journal of The Royal Society Interface*, vol. 17, no. 170, p. 20200482, Sep. 2020.

[352] V. Dakos and F. Soler-Toscano, "Measuring complexity to infer changes in the dynamics of ecological systems under stress," *Ecological Complexity*, vol. 32, pp. 144–155, Sep. 2016.

[353] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed., ser. Springer Series in Synergetics. Berlin: Springer, 2009.

Appendix A

Supplementary Material for Chapter 2

Table A.1: Expanded glossary of key concepts underpinning tipping point theory, explained in terms of their mathematical and intuitive/ecological definitions.

Concept	Mathematical Definition		Intuitive or Ecological Definition
	In univariate (1D) systems	***In higher dimensions***	
Dynamical System	$\frac{dx}{dt} = f(x, B)$	Same but in vector / matrix notation	A system that changes over time in a manner defined by the growth function $f(x, B)$, which depends on the current state of the system, x, and the system parameter(s) B (i.e., conditions, such as climate or environmental conditions like precipitation, temperature etc.).
Stochastic Dynamical System	$dx = f(x, B)dt + g(x, B)dW_t$ Where: f is the deterministic ('drift') term, g is the stochastic ('diffusion') term.	Same but in vector / matrix notation	System with certain degree of randomness such that it is ever-fluctuating about its current stable state. Most real natural systems are understood to be stochastic due to variability in their forcing (changing of the system parameter(s)) or random internal noise (captured here by dW_t).
Equilibrium	A state x^* where $f(x^*, B) = 0$		Values of the state variable where the growth function is zero. As a consequence, if the system is at an equilibrium and is unperturbed, it will remain at this equilibrium.
Stability	Determined by: $\left.\frac{df(x, B)}{dx}\right\|_{x=x^*}$ Gradient of $f(x, B)$ w.r.t. x, evaluated at equilibrium x^*. Can be denoted $f'(x^*, B)$ or, often, λ. $\lambda < 0 \rightarrow$ stable, $\lambda > 0 \rightarrow$ unstable.	Determined by the dominant eigenvalue (eigenvalue with largest real part) of the Jacobian matrix (matrix of first order partial derivatives) of $f(x, B)$. The dominant eigenvalue, also denoted λ, is a measure of stability along the direction in state space that is least stable. $\mathrm{Re}\lambda < 0 \rightarrow$ stable, $\mathrm{Re}\lambda > 0 \rightarrow$ unstable.	Stability refers to the behaviour of a system following small perturbations from an equilibrium. An equilibrium point is said to be stable if trajectories return to equilibrium following perturbation. In contrast, trajectories move away from unstable states when perturbed. The dominant eigenvalue, λ (which is the local gradient at equilibrium in univariate systems), gives a measure for the rate at which the system returns to an equilibrium (if stable) or moves away from it (if unstable).

Concept	Mathematical Definition		Intuitive or Ecological Definition
	In univariate (1D) systems	***In higher dimensions***	
Characteristic Return Time	$T_R = -1/\lambda$	$T_R = -1/\mathrm{Re}\lambda$	Provides a time scale that quantifies how quickly the system returns to equilibrium.
	As a system returns to equilibrium following a perturbation, its distance from equilibrium reduces by a factor of e after a time T_R [89].		
Critical Slowing Down	The dominant eigenvalue approaching zero from below, due to changing parameters. This corresponds to the characteristic return time approaching infinity.		System responds more slowly to perturbations following a loss in resilience decreasing system stability, e.g., forest takes longer to recover from drought or fire.
Potential Landscape	The potential of a system is defined: $$V(x) = -\int f(x, B)dx$$ Note that this will have stationary points (peaks and valleys) when $f(x, B) = 0$, i.e. at x^*, but the negative sign makes the stable equilibria into valleys and unstable ones into peaks.		While the concept of potential is somewhat arbitrary in ecological terms (square of system state per unit time), it serves as a useful mathematical and conceptual model for the stability and resilience of a system, with the valleys at stable equilibria corresponding to '**basins of attraction**' [97] separated by peaks from unstable equilibria. This is also sometimes described in terms of the '**ball-and-cup**' model wherein the system state corresponds to a ball that moves along the potential landscape (with no inertia) and the basins of attraction are the 'cups'. At a stable equilibrium, the ball is at the bottom of one of the cups but is constantly moving from small perturbations due to stochasticity (randomness) in external forcing and random internal noise. However, the walls of the 'cup' cause it to constantly roll back to equilibrium, at a rate depending on the steepness of the walls – which we can see from the mathematics relates to λ. The amount of disturbance that the system can withstand is described by the shape and size of the basin of attraction [95]. Specifically, the width indicates how much the system state variable can be changed before it leaves the current basin of attraction, while the depth and steepness relate to how much the forcing and noise terms can change before the system switches to an alternative attractor.

Concept	Mathematical Definition		Intuitive or Ecological Definition
	In univariate (1D) systems	***In higher dimensions***	
	If the underlying dynamics of the system are not known, one option to estimate the potential landscape is to integrate the Taylor expansion of the growth function at some small perturbation about equilibrium, $f(x^* + \eta)$ (as in e.g., [187]): $$\frac{d(x^* + \eta)}{dt} = f(x^* + \eta)$$ $$= f(x^*) + \left.\frac{\partial f}{\partial x}\right|_{x^*} \eta + \frac{1}{2}\left.\frac{\partial^2 f}{\partial x^2}\right|_{x^*} \eta^2 + \dots$$ $$f(x^* + \eta) = \lambda_1 \eta + \lambda_2 \eta^2 + \dots$$ $$V(\eta) = -\int f(x^* + \eta) d\eta = -\frac{1}{2}\lambda_1 \eta^2 - \frac{1}{3}\lambda_2 \eta^3 + \dots$$		If we only take the Taylor expansion of $f(x^* + \eta)$ to the first-order term (a.k.a. linearising the equation), the resulting potential landscape will only have a quadratic term corresponding to a parabola. This is a reasonable approximation of the potential landscape near the equilibrium (capturing the dynamics for small perturbations) and will get shallower as λ_1 (the dominant eigenvalue) approaches zero. However, it ignores other stable equilibria, which require higher-order terms. Adding a cubic term will add a 'peak' roughly corresponding to the unstable equilibrium, but a bi-stable system will require at least a fourth-order term. Even with just the cubic approximation, though, as λ_1 approaches zero the peak and valley approach and eventually coincide, becoming a flat point of inflection leaving the system to fall into a neighbouring basin of attraction.
Resilience	A wide range of definitions and associated metrics of resilience have been proposed, though the two most prevailing were put forward by Holling [90, 91] and are termed '**engineering resilience**' and '**ecological resilience**'.		
	Engineering resilience essentially corresponds to stability and the characteristic return time from perturbations, T_R.		How fast a system recovers after perturbation [90, 91, 218].
	Ecological resilience is generally characterised by the shape and size of the basin of attraction – a number of metrics related to the features of it and the potential landscape are outlined in [95].		The amount of disturbance, perturbation or forcing an ecosystem can endure without changing its function and passing to an alternative stable state. This can be facilitated by mechanisms that are not necessarily fully captured by dynamical systems theory, such as the capacity of ecosystems to reorganise, though such changes should result in an altered potential landscape and basin of attraction.
Bifurcations and Normal Forms	In dynamical systems theory, a bifurcation is a qualitative, topological change in the dynamics of a system in response to gradual, smooth change in the system parameter(s) – i.e., condition(s) passing a critical threshold [69, 79]. For bifurcation-induced tipping points, this change is typically the loss of the current equilibrium state's stability, resulting in a transition to an alternate stable state. Bifurcations can be grouped into basic, lower-dimensional 'normal forms' which characterise different responses to forcing and to which the dynamics of high dimensional systems simplify near a bifurcation (**centre manifold theorem**). Three of the most prevalent in ecology are:		

Concept	Mathematical Definition		Intuitive or Ecological Definition
	In univariate (1D) systems	*In higher dimensions*	
	Transcritical bifurcation – continuous, gradual transition from one stable state to another. This corresponds to a 'kink' in the **bifurcation diagram** (which plots the system's equilibrium values against the system parameter(s); Fig. 2.1e,f Main Text) owing to the second order derivative of the equilibrium values with respect to the system parameter(s) (i.e., of the bifurcation diagram's curve; $\frac{d^2x^*}{dB^2}$) responding discontinuously to forcing [79, 105].		This can characterise "the transition to population extinction" [352, p. 145]. Some of earliest transcritical bifurcations in ecology were effectively first described by Lotka in 1925 [96].
	Hopf bifurcation – a transition from a single stable equilibrium to "stable cyclical dynamics" [105, p. 906].		First described in ecology in terms of "the onset of oscillations in predator-prey models" [352, p. 145] by Rosenzweig and Macarthur in 1963 [100].
	Fold (or sometimes, "catastrophic" or "saddle node") **bifurcation** – characterised by an abrupt and discontinuous shift to an alternative equilibrium [105, p. 906]. It is the result of a stable and an unstable equilibrium colliding and annihilating one-another (Fig. 2.1b,d,f Main Text).		Fold bifurcations typify tipping points in dynamical systems given they are particularly abrupt and difficult (sometimes impossible) to reverse, owing to **hysteresis**: the need to return to a lower parameter threshold than that which brought about the new state in order to revert to the previous stable state. The name refers to the 'fold' in the curve of the bifurcation diagram (Fig. 2.1f, Main Text). Fold bifurcations were first described in ecology in terms of "resource overexploitation" [352, p. 145] in Noy-Meir's [99] 1975 extension of graphical analyses of predator-prey models by Rosenzweig and Macarthur [100] to grazing systems. They were further discussed in relation to various classic ecological models (including for harvesting animal populations and pest outbreaks) in a seminal review paper by Robert May in 1977 [98].
Tipping Points	The point at which small change can make a big difference to a system [59, 75]. Specifically, where a small perturbation or increment in the system parameter(s) (e.g., climate and environmental conditions) causes a qualitative change to the system driven by strong positive feedback, such as a transition to an alternate stable state. Use of the term has varied in some previous literature [76]. As it is typified by fold bifurcations (but can describe other phenomena that produce state transitions or other qualitative change), it is sometimes interpreted as being synonymous with the bifurcation point. However, tipping points can be caused by noise or small perturbations near the bifurcation point, which may be sufficient to push the system out of its current basin of attraction to an alternate stable state.		

Concept	Mathematical Definition	Intuitive or Ecological Definition
	In univariate (1D) systems	
Early Warning Signal - Autocorrelation	In continuous time, the autocorrelation of the system state fluctuations is given by: $$\rho(\tau) = e^{\lambda\tau}$$ It thus tends to one as $\lambda \to 0$ from below. In discrete time, if the system can be well described by an autoregressive process of order 1 (AR(1)) then the lag-1 autocorrelation is equivalent to the AR(1) coefficient: $$\alpha = e^{\lambda\Delta t}$$ Thus the lag-1 autocorrelation and AR(1) coefficient also tend to 1 for processes best modelled in discrete time as $\lambda \to 0$ from below. Full derivations are provided below. Please note that the equations here are for 1D systems and become more complex in higher dimensions.	Critical slowing down causes the system's current state to be more similar to its recent previous states (sometimes described conceptually in terms of rising memory of the system [80, 105]), hence there is an increase in the system state time series of autocorrelation at lag-1. This also holds true for further lags for transcritical and fold bifurcations, but is not necessarily the case for all types of bifurcation, such as Hopf bifurcations, for which autocorrelation may increase or decrease at subsequent lags [351].
Early Warning Signal - Variance	In continuous time, the variance of the system state fluctuations is given in the stationary regime ($t \to \infty$, assuming $\lambda < 0$) by: $$\mathrm{Var}(\eta) = -\frac{\sigma^2}{2\lambda}$$ Where $\eta = x - x^*$. The variance thus tends to infinity as $\lambda \to 0$ from below. In discrete time, the variance is given by: $$\mathrm{Var}(\eta) = \frac{\sigma^2}{1-\alpha^2}$$ Where α is the AR(1) coefficient introduced in the autocorrelation section above. Since $\alpha \to 1$ as $\lambda \to 0$, the discrete time variance also tends to infinity under critical slowing down. Full derivations are also provided below. As above, these equations are for 1D systems and are more complex in higher dimensions.	As critical slowing down produces longer and larger fluctuations in the system state due to the system's declining capacity to recover from perturbations, variance of the system state and its fluctuations increase over time.

A.1 Derivation of variance and autocorrelation as tipping point EWS

A.1.1 Continuous-time process

A general stochastic process with additive white noise may be written

$$\frac{dx}{dt} = f(x) + \sigma\xi(t) \tag{A.1}$$

where f describes the deterministic dynamics, σ is the noise amplitude, and $\xi(t)$ is a Gaussian white noise process satisfying $\mathbb{E}[\xi(t)] = 0$ and $\mathbb{E}[\xi(t)\xi(s)] = \delta(t-s)$. Assuming a stable equilibrium at x^*, and letting $\eta(t) = x - x^*$, we have

$$\frac{d\eta}{dt} = f(x^* + \eta) + \sigma\xi(t), \tag{A.2}$$

an equation which describes the dynamics about equilibrium. For sufficiently small deviations from equilibrium, the dynamics are well approximated by low-order terms in the Taylor expansion. Expanding $f(x^* + \eta)$ in a Taylor series about x^* and retaining only the linear term (valid for small η),

$$\frac{d\eta}{dt} = \lambda\eta + \sigma\xi(t), \tag{A.3}$$

where $\lambda = \frac{df}{dx}|_{x=x^*}$. This equation is known as an Ornstein Uhlenbeck Process. Its statistical properties are well known (e.g. Gardiner's *Handbook of stochastic methods* [353]). It can be solved using an integrating factor $e^{-\lambda t}$, and has solution

$$\eta(t) = \eta(0)e^{\lambda t} + \sigma e^{\lambda t}\int_0^t e^{-\lambda t'}\xi(t')dt'. \tag{A.4}$$

In the long-time limit ($t \to \infty$), the contribution from the initial condition $\eta(0)e^{\lambda t}$ vanishes since $\lambda < 0$ (the equilibrium is stable).

Variance

The variance is defined as

$$\mathrm{Var}(\eta) = \mathbb{E}[(\eta - \mathbb{E}[\eta])^2] \tag{A.5}$$

Since $\mathbb{E}[\xi(t)] = 0$ and the process is linear in ξ, we have $\mathbb{E}[\eta(t)] = 0$, so

$$\mathrm{Var}(\eta) = \mathbb{E}[\eta^2] \tag{A.6}$$

Substituting the solution for $\eta(t)$ in the long-time regime (where the deterministic term vanishes), we get

$$\mathrm{Var}(\eta) = \mathbb{E}\left[\left(\sigma e^{\lambda t}\int_0^t e^{-\lambda t'}\xi(t'), dt'\right)^2\right] \tag{A.7}$$

$$= \sigma^2 e^{2\lambda t}\int_0^t dt' \int_0^t dt'', e^{-\lambda(t'+t'')}\mathbb{E}[\xi(t')\xi(t'')]. \tag{A.8}$$

Using the white noise property $\mathbb{E}[\xi(t')\xi(t'')] = \delta(t' - t'')$, we find

$$\mathrm{Var}(\eta) = \sigma^2 e^{2\lambda t}\int_0^t dt', e^{-2\lambda t'} \tag{A.9}$$

$$= \sigma^2 e^{2\lambda t}\cdot\frac{1 - e^{-2\lambda t}}{2\lambda} \tag{A.10}$$

$$= \frac{\sigma^2}{2\lambda}\left(e^{2\lambda t} - 1\right), \tag{A.11}$$

which simplifies in the stationary regime ($t \to \infty$, assuming $\lambda < 0$) to

$$\mathrm{Var}(\eta) = -\frac{\sigma^2}{2\lambda}. \tag{A.12}$$

Note that as $\lambda \to 0$ from below, the variance diverges, indicating a loss of stability near bifurcation (i.e., critical slowing down).

Autocorrelation

The autocorrelation function is defined as

$$\rho(\tau) = \frac{\mathbb{E}[\eta(t)\eta(t+\tau)]}{\mathrm{Var}(\eta)}. \tag{A.13}$$

The numerator is given by

$$\mathbb{E}[\eta(t+\tau)\eta(t)] = \mathbb{E}\left[\left(\sigma e^{\lambda(t+\tau)}\int_0^{t+\tau} e^{-\lambda t''}\xi(t'')dt''\right)\left(\sigma e^{\lambda t}\int_0^{t} e^{-\lambda t'}\xi(t')dt'\right)\right] \tag{A.14}$$

$$= \sigma^2 e^{2\lambda t}e^{\lambda\tau}\int_0^{t+\tau} dt''\int_0^{t} dt' e^{-\lambda(t''+t')}\delta(t'-t'') \tag{A.15}$$

$$= \sigma^2 e^{2\lambda t}e^{\lambda\tau}\int_0^{t+\tau} dt'' e^{-2\lambda t''} \tag{A.16}$$

$$= \frac{\sigma^2 e^{\lambda\tau}}{2\lambda}\left(e^{2\lambda t}-1\right), \tag{A.17}$$

which, in the stationary regime ($t \to \infty$) gives an autocorrelation function of

$$\rho(\tau) = e^{\lambda\tau}. \tag{A.18}$$

Note that as $\lambda \to 0$ from below, the autocorrelation tends to 1, another indication of critical slowing down.

A.1.2 Discrete-time process

Some systems are more naturally modeled in discrete time (such as population dynamics with non-overlapping generations). Discrete-time processes also arise as numerical approximations to continuous-time systems when the time step is small.

A general form of a discrete-time stochastic process with additive white noise is

$$x_{t+1} = f(x_t) + \sigma\epsilon_t, \tag{A.19}$$

where f describes the deterministic dynamics, σ is the noise amplitude, and $\epsilon_t \sim \mathcal{N}(0,1)$ are independent and identically distributed (i.i.d.) standard Normal random variables.

Let x^* be a fixed point of the deterministic dynamics, i.e., an equilibrium satisfying $f(x^*) = x^*$. To study fluctuations about equilibrium, define the residual $\eta_t = x_t - x^*$, so that the dynamics are centered around x^*. Then

$$\eta_{t+1} = f(x^* + \eta_t) - x^* + \sigma\epsilon_t \tag{A.20}$$

Taking a first-order Taylor expansion of f about x^* and noting that $f(x^*) = x^*$, we have

$$f(x^* + \eta_t) \approx f(x^*) + f'(x^*)\eta_t = x^* + \alpha\eta_t,$$

where $\alpha = f'(x^*)$. Substituting into the update rule yields

$$\eta_{t+1} = \alpha\eta_t + \sigma\epsilon_t. \tag{A.21}$$

This is known as an autoregressive process of order 1 (AR(1)). It is the discrete-time analogue of the Ornstein–Uhlenbeck process.

Variance

Taking the expectation of both sides of the AR(1) equation,

$$\mathbb{E}[\eta_{t+1}] = \alpha\mathbb{E}[\eta_t] + \sigma\mathbb{E}[\epsilon_t], \tag{A.22}$$

and using $\mathbb{E}[\epsilon_t] = 0$, we find that the mean satisfies

$$\mathbb{E}[\eta_{t+1}] = \alpha\mathbb{E}[\eta_t].$$

Assuming stationarity, the mean is constant in time, so $\mathbb{E}[\eta_{t+1}] = \mathbb{E}[\eta_t] = \mu$. This gives

$$\mu = \alpha\mu \quad \Rightarrow \quad \mu(1-\alpha) = 0,$$

implying $\mu = 0$ (for $\alpha \neq 1$). Squaring both sides of the AR(1) equation and taking expectations yields

$$\mathbb{E}[\eta_{t+1}^2] = \alpha^2\mathbb{E}[\eta_t^2] + 2\alpha\sigma\mathbb{E}[\eta_t\epsilon_t] + \sigma^2\mathbb{E}[\epsilon_t^2]. \tag{A.23}$$

Since η_t and ϵ_t are independent and $\mathbb{E}[\epsilon_t] = 0$, we have $\mathbb{E}[\eta_t\epsilon_t] = 0$, and $\mathbb{E}[\epsilon_t^2] = 1$. Thus,

$$\mathbb{E}[\eta_{t+1}^2] = \alpha^2\mathbb{E}[\eta_t^2] + \sigma^2. \tag{A.24}$$

Assuming stationarity (i.e., constant variance), we get

$$\mathrm{Var}(\eta) = \alpha^2\mathrm{Var}(\eta) + \sigma^2, \tag{A.25}$$

which rearranges to

$$\mathrm{Var}(\eta) = \frac{\sigma^2}{1-\alpha^2}. \tag{A.26}$$

This expression is valid only for $|\alpha| < 1$, which ensures that the process is stationary and the variance remains finite. The variance diverges as $\alpha \to 1$ from below, an indication of critical slowing down. This is the expression given by Scheffer et al. in [74].

Lag-1 autocorrelation

Multiplying the AR(1) equation by a factor of η_t, and taking expectations gives

$$\mathbb{E}[\eta_{t+1}\eta_t] = \alpha\mathbb{E}[\eta_t^2] + \sigma\mathbb{E}[\epsilon_t\eta_t]. \tag{A.27}$$

Since η_t and ϵ_t are independent and have zero mean, $\mathbb{E}[\epsilon_t\eta_t] = 0$. Thus,

$$\mathbb{E}[\eta_{t+1}\eta_t] = \alpha\mathbb{E}[\eta_t^2]. \tag{A.28}$$

Dividing both sides by $\mathbb{E}[\eta_t^2]$, we obtain the lag-1 autocorrelation:

$$\rho(1) = \frac{\mathbb{E}[\eta_{t+1}\eta_t]}{\mathbb{E}[\eta_t^2]} = \alpha. \tag{A.29}$$

The lag-1 autocorrelation tends to one as $\alpha \to 1$ from below, another indication of critical slowing down. Note that here, the AR(1) coefficient (α) is equal to the lag-1 autocorrelation. This is not the case in general (as a stochastic process may not be well described by an AR(1) process in general). However, systems at equilibrium subject to small perturbations are well described by an AR(1) process, hence the AR(1) coefficient and lag-1 autocorrelation measured in data should be similar.

Appendix B

Supplementary Material for Chapter 3

Table B.1: Breakdown of *Zylia* ZM-1 and MAARU recordings used for each Sub-Project of this study, as well as the ecoacoustic metrics extracted from MAARU's recordings. The ZM-1 recordings for Sub-Project 3 omit Site 6, as there was insufficient space about the microphone to record the playback of sinusoidal sweeps that were used to generate the impulse response of reflections at four points about the microphone and subsequently simulate Site 6's reverberation for added bird calls.

	Sub-Project 1 (Efficacy of Soundscape Reproduction)	**Sub-Project 2 (Impact of Device Orientation)**	**Sub-Project 3 (BirdNET and HARKBird Performance on Additional Avian Calls)**
ZM-1 3rd Order Ambisonic Soundscape Recordings Used	6 sites x 10 min. recordings *(1 recording for each of 6 sites)*	6 sites x 10 min. recordings *(1 recording for each of 6 sites)*	5 sites x 10 min. recordings *(1 for each of first 5 sites)* with added bird calls *(1 species during a quieter section of each recording — see Table 5)* x 2 methods of adding bird calls *(encoding to ambisonics domain and playback from individual loudspeakers)*
MAARU Recordings Used	6 sites x 10 min. recordings x 2 recording conditions *(field recordings and recordings in vertical orientation of ZM-1-captured soundscapes reproduced through VSE)*	Sub-Project 1 Recordings + 2 additional VSE recording conditions *(recordings at 45° pitch angle and in horizontal orientation of ZM-1-captured soundscapes reproduced through VSE)*	Recordings in horizontal orientation of ZM-1-captured soundscapes with added bird calls reproduced through VSE
Total Duration of ZM-1 Recordings Used	60 mins		100 mins
Total Duration of MAARU Recordings Used	240 mins *(6 sites x 10 min. recordings x 4 recording conditions total)*		100 mins *(5 sites x 10 min. recordings x 2 methods of adding bird calls)*
Metrics Extracted from MAARU Recordings	Monophonic (extracted from upper left microphone of array): Power spectrograms and spectral differences, 7 Acoustic Indices and VGGish feature embedding, and BirdNET classifications Multichannel (extracted from full 6-channel recordings): HARKBird localisations		Monophonic (extracted from upper left microphone of array): BirdNET classifications Multichannel (extracted from full 6-channel recordings): HARKBird localisations

Table B.2: BirdNET's Precision (Prec.), recall (Rec.), number of predicted species (Spec.) and number of predictions (Pred.) across all recording sites and conditions (field and Virtual Sound Environment (VSE) at various orientations) towards Sub-Projects 1 (Efficacy of Soundscape Reproduction) and 2 (Impact of Device Orientation) for the following minimum confidence thresholds (leftmost column): 0, 0.25, 0.5, 0.75 and 0.9.

Condition		Field						VSE - Vertical						VSE - 45°						VSE - Horiztonal					
Site		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
0	Prec.	0.85	0.93	-	0.50	0.81	0.78	0.64	0.76	0	0	0.87	0.54	0.50	0.90	-	1	0.67	0.53	0.86	0.83	0	0.50	0.70	0.6
	Rec.	0.08	0.18	0	0.02	0.24	0.30	0.17	0.25	0	0	0.31	0.21	0.10	0.19	0	0.02	0.29	0.13	0.26	0.26	0	0.02	0.31	0.27
	Spec.	2	6	0	2	5	5	11	7	2	1	5	9	9	4	0	1	10	6	5	4	1	2	10	9
	Pred.	13	27	0	2	32	27	36	45	2	1	38	26	26	29	0	1	45	17	42	42	1	2	46	30
0.25	Prec.	1	1	-	-	0.94	0.94	0.82	0.83	-	-	0.96	0.63	0.69	0.92	-	-	0.87	0.75	0.97	0.94	-	-	0.83	0.71
	Rec.	0.04	0.13	0	0	0.16	0.21	0.10	0.22	0	0	0.24	0.07	0.06	0.17	0	0	0.25	0.04	0.20	0.23	0	0	0.28	0.14
	Spec.	1	3	0	0	2	2	4	4	0	0	2	3	5	3	0	0	4	3	3	3	0	0	6	4
	Pred.	6	18	0	0	18	16	17	35	0	0	27	8	13	25	0	0	31	4	30	34	0	0	35	14
0.5	Prec.	1	1	-	-	0.92	1	1	0.96	-	-	0.95	0.67	0.86	0.95	-	-	0.90	-	1	0.93	-	-	0.95	0.67
	Rec.	0.02	0.10	0	0	0.10	0.03	0.06	0.17	0	0	0.17	0.03	0.04	0.13	0	0	0.18	0	0.13	0.20	0	0	0.19	0.06
	Spec.	1	2	0	0	2	1	1	2	0	0	2	2	3	2	0	0	2	0	1	1	0	0	4	2
	Pred.	3	14	0	0	12	2	9	25	0	0	20	3	7	19	0	0	21	0	18	29	0	0	22	6
0.75	Prec.	1	1	-	-	0.75	1	1	1	-	-	1	1	1	1	-	-	0.86	-	1	1	-	-	1	1
	Rec.	0.01	0.08	0	0	0.03	0.01	0.03	0.13	0	0	0.06	0.01	0.02	0.10	0	0	0.06	0	0.08	0.14	0	0	0.09	0.03
	Spec.	1	2	0	0	2	1	1	1	0	0	1	1	1	2	0	0	2	0	1	1	0	0	3	1
	Pred.	2	11	0	0	4	1	5	18	0	0	7	1	3	14	0	0	7	0	11	20	0	0	10	2
0.9	Prec.	-	1	-	-	-	-	1	1	-	-	1	-	1	1	-	-	1	-	1	1	-	-	1	-
	Rec.	0	0.07	0	0	0	0	0.02	0.09	0	0	0.02	0	0.02	0.08	0	0	0.01	0	0.03	0.09	0	0	0.01	0
	Spec.	0	2	0	0	0	0	1	1	0	0	1	0	1	1	0	0	1	0	1	1	0	0	1	0
	Pred.	0	9	0	0	0	0	3	12	0	0	2	0	3	11	0	0	1	0	4	13	0	0	1	0

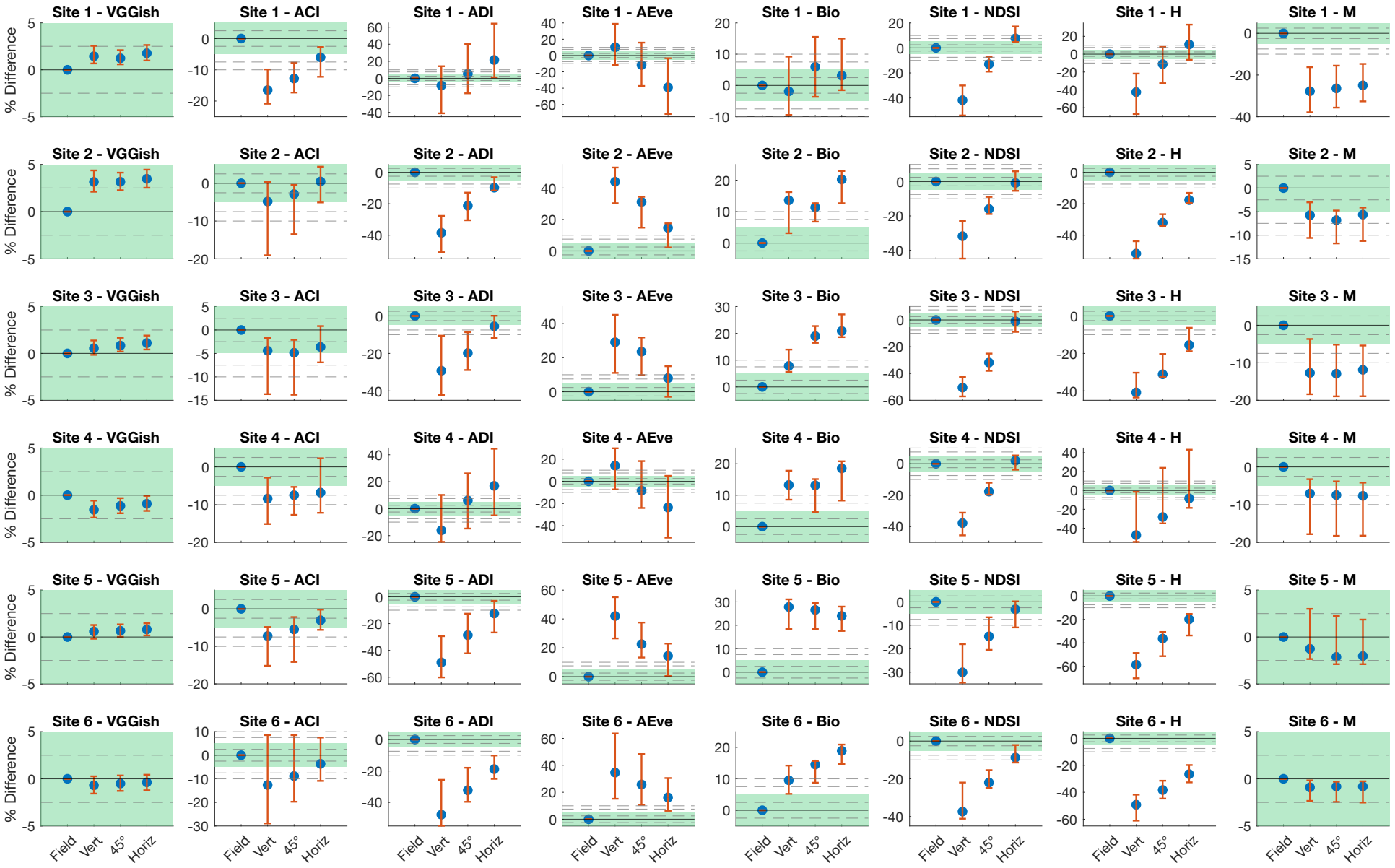


Figure B.1: Plots of scaled differences in indices' values between VSE and field recordings as a percentage of each index's field recording range, for all sites. Blue dots indicate the median percentage differences, whiskers show the upper and lower quartiles, the shaded area denotes the ±5% limits of agreement and, where appropriate, the grey dashed lines represent ±2.5%, ±7.5% and ±10% as potential alternative limits agreement.

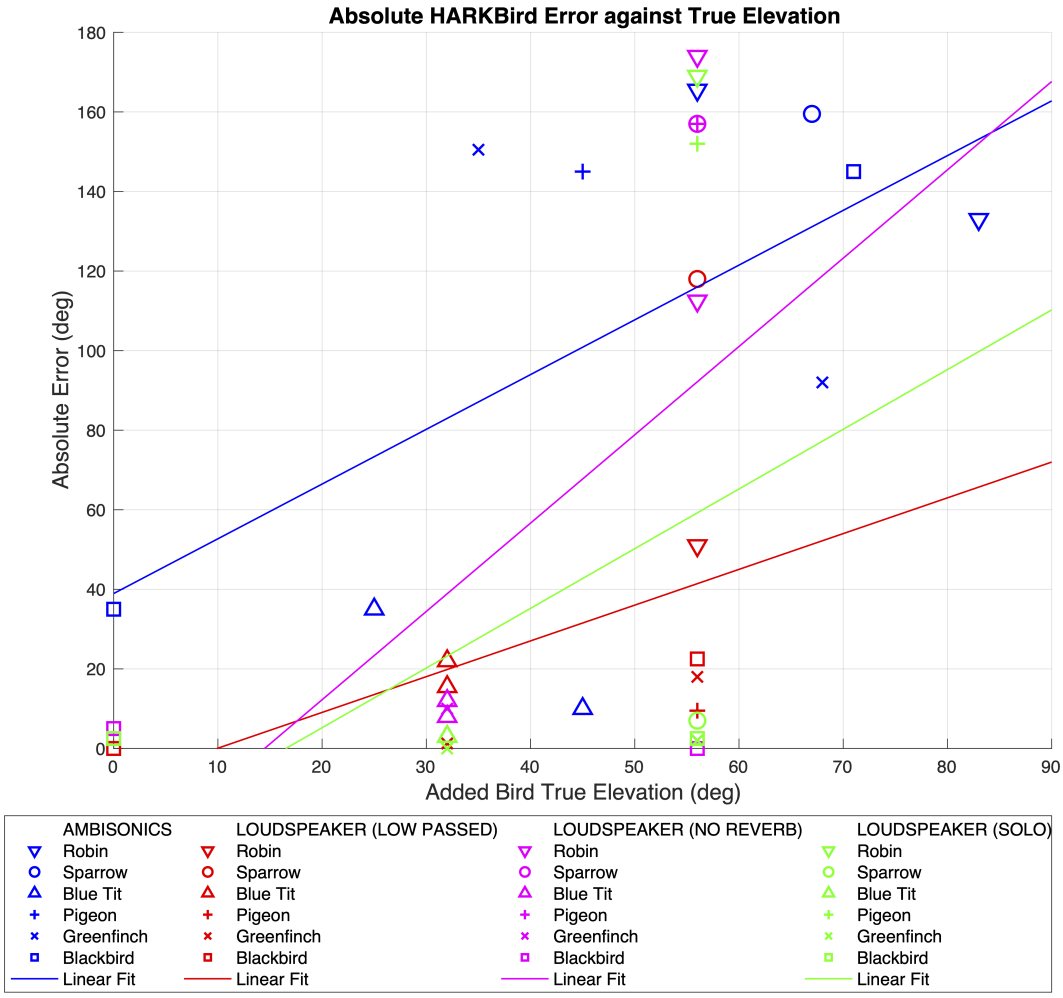


Figure B.2: Absolute error in HARKBird's estimates of the azimuth of avian calls embedded in VSE-replicated soundscapes against calls' true elevation values, for four methods of embedding: ambisonic encoding, playback from individual loudspeakers (recordings low-passed with a 4 kHz cut-off and 12 dB roll-off prior to passing to HARKBird), individual loudspeaker playback with no reverb, and soloed playback from individual loudspeakers. In all cases errors are larger at higher elevation values, as these calls are further from the plane in which all of the recording device's microphones are located. Performance is best, however, from low-passed recordings of calls played back from individual loudspeakers.

Appendix C

Supplementary Material for Chapter 4

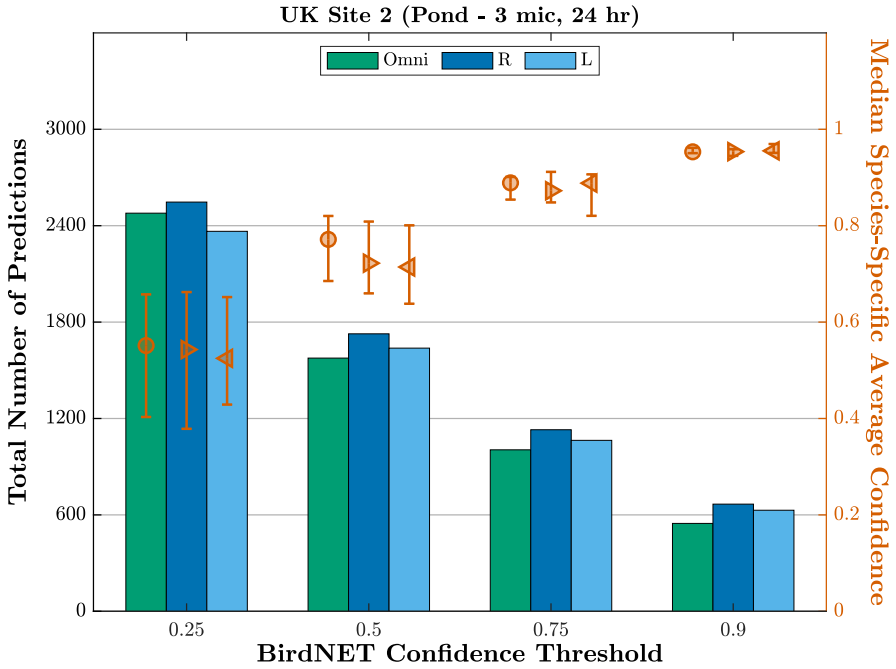


Figure C.1: Version of Fig. 4.5 for UK Site 2 (Pond) using three microphones for beamforming (configured as shown in Fig. 4.2). Abbreviations for beamforming direction: Omni - Omnidirectional; R - Right; L - Left.

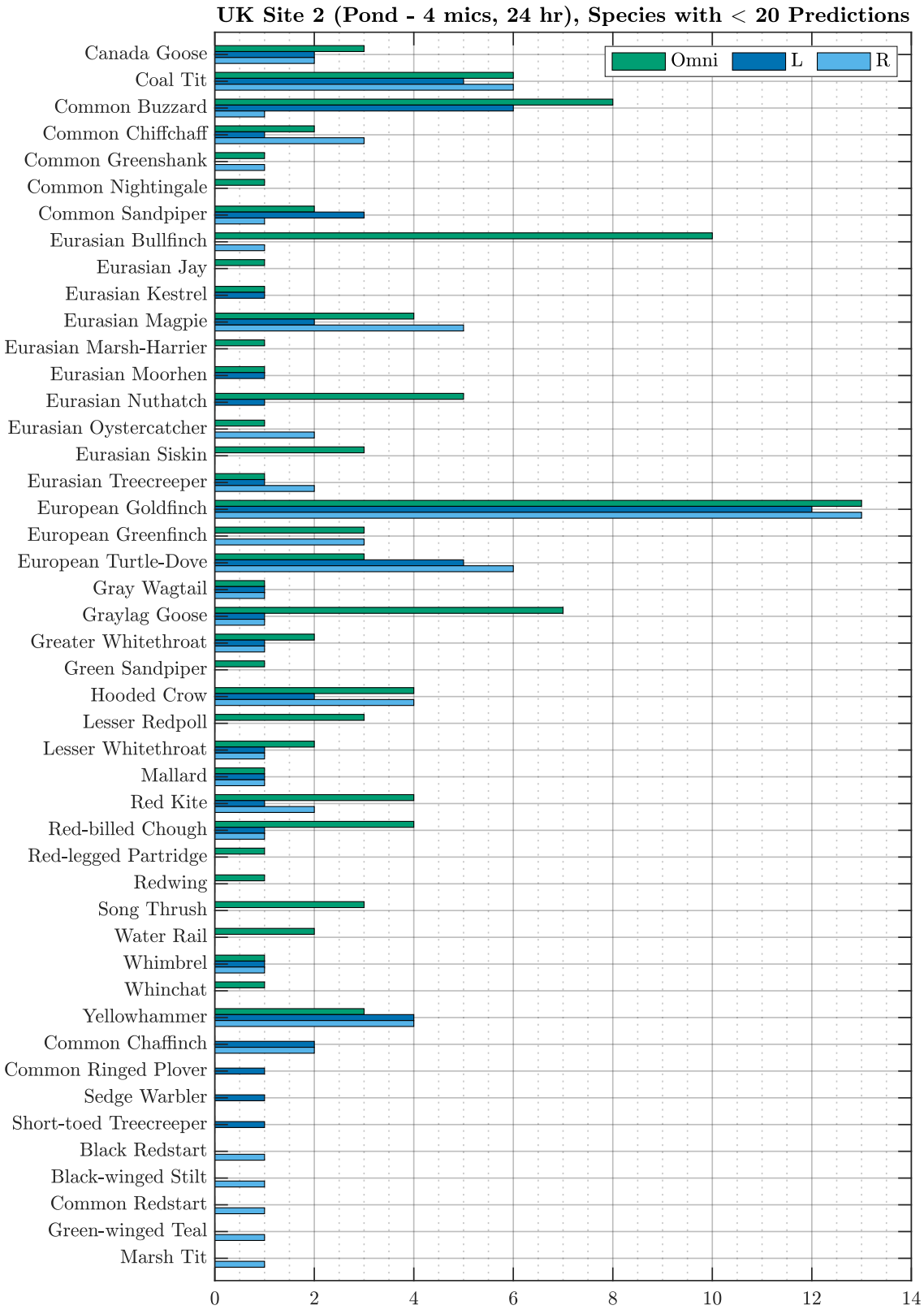


Figure C.2: Version of Fig. 4.7 showing remaining species for UK Site 2 (Pond). Note that here, we count all species predicted by BirdNET for each recording condition, not just common species that appear in all recording conditions as in earlier analyses. Abbreviations for beamforming direction: Omni - Omnidirectional; R - Right; L - Left.

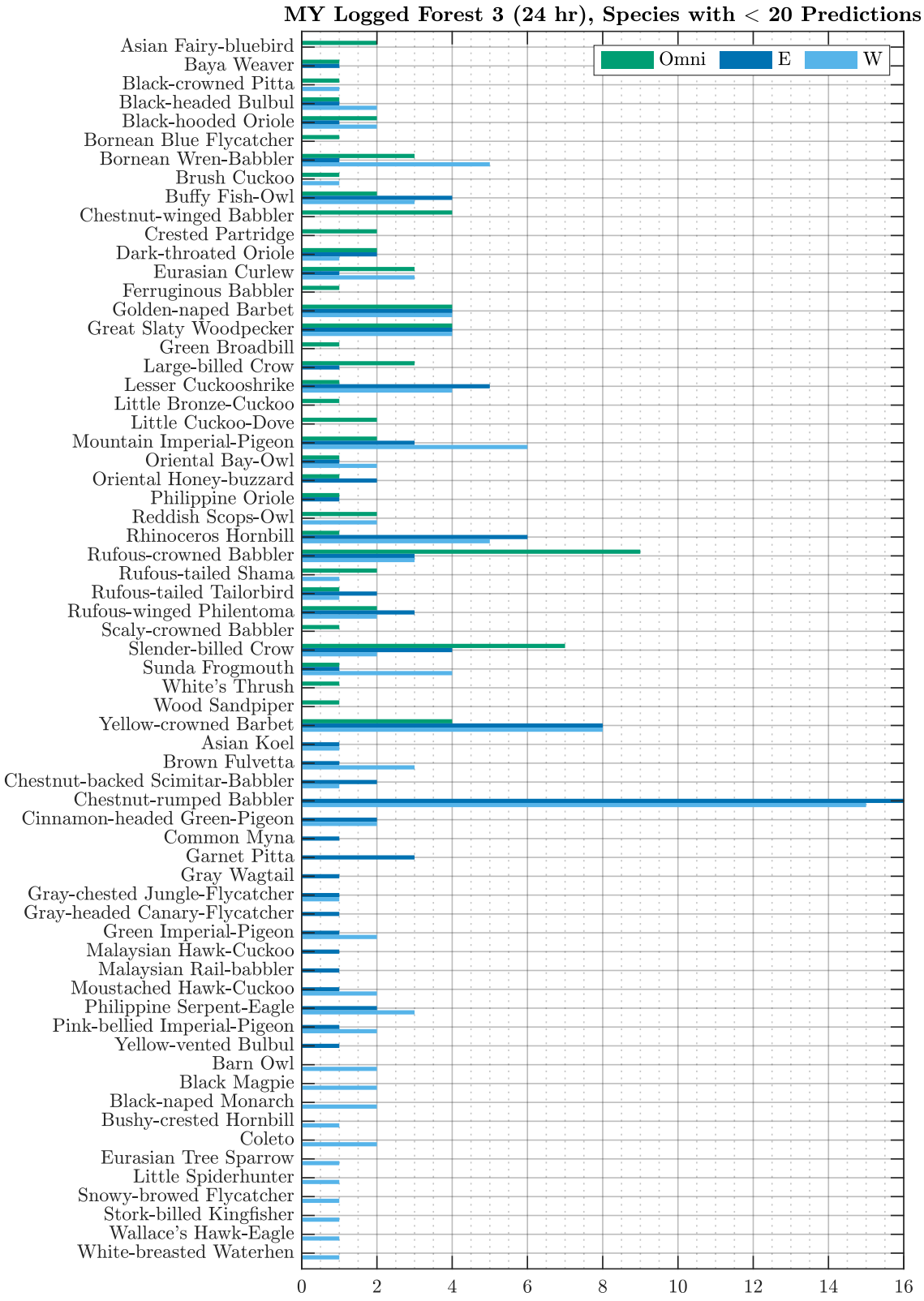


Figure C.3: Version of Fig. 4.7 showing remaining species for Malaysia Logged Forest 3. Note that here, we count all species predicted by BirdNET for each recording condition, not just common species that appear in all recording conditions as in earlier analyses. Abbreviations for beamforming direction: Omni - Omnidirectional; E - East; W - West.

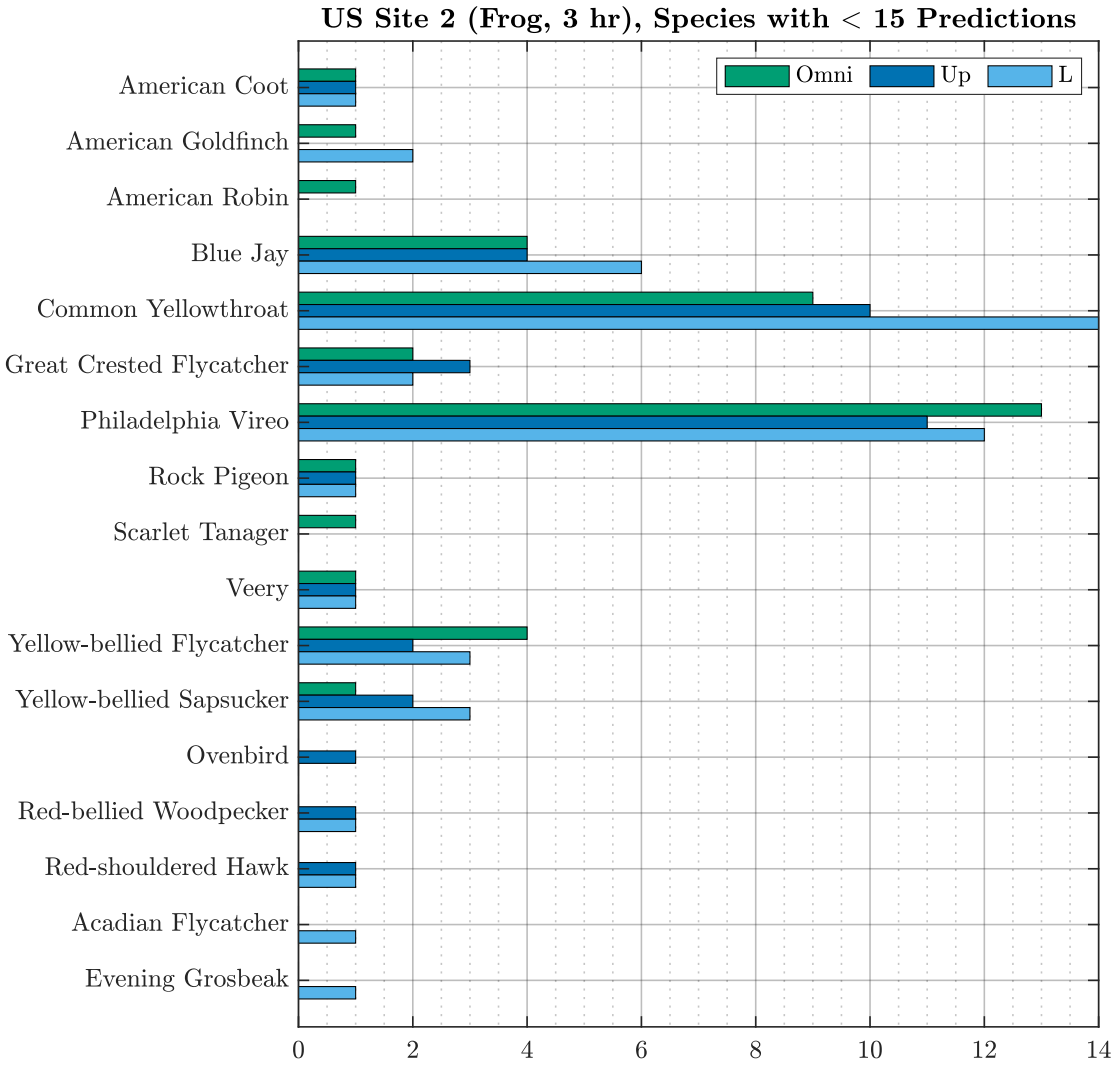


Figure C.4: Version of Fig. 4.7 showing remaining species for US Site 2 (Frog). Note that here, we count all species predicted by BirdNET for each recording condition, not just common species that appear in all recording conditions as in earlier analyses. Abbreviations for beamforming direction: Omni - Omnidirectional; L - Left-downwards (i.e., $-135°$ azimuth).

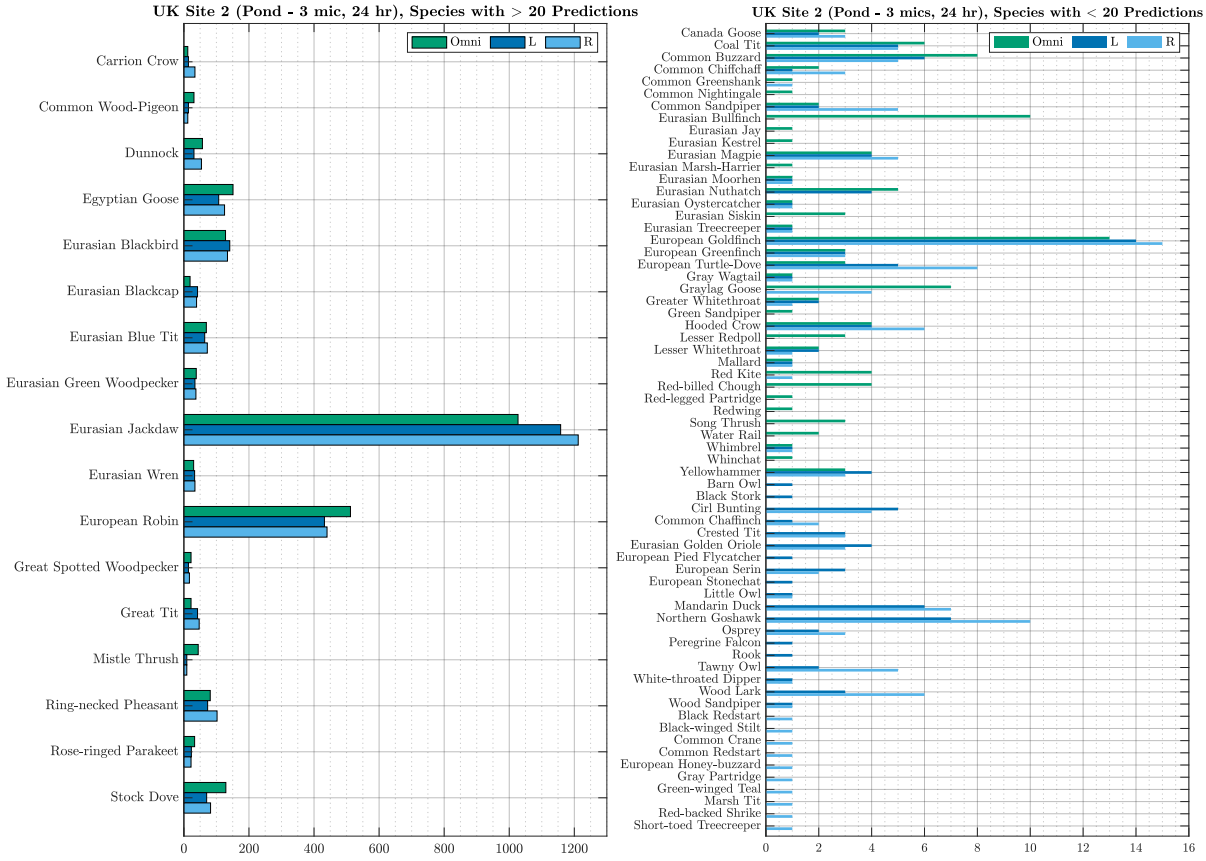


Figure C.5: Number of BirdNET predictions above confidence 0.25 for all species in omnidirectional audio and audio beamformed with the subset of three microphones shown in Fig. 4.2 at UK Site 2 (Pond), where beamforming was used to attempt to distinguish sites of distinct neighbouring land use. Abbreviations for beamforming direction: Omni - Omnidirectional; R - Right; L - Left.

Appendix D

Supplementary Material for Chapter 5

Table D.1: Values of slope and breakpoint metrics of satellite-sensed vegetation resilience in Borneo and the Amazon.

(a) Values of pre- and/or post-breakpoint slope of spatially-averaged VOD AC(1) time series for Borneo and the Amazon, and average pre- and post-breakpoint slope for individual grid cells' VOD AC(1). Abbreviations: AC(1): lag-1 autocorrelation; CrI: Credible Interval; SD: Standard Deviation; VOD: Vegetation Optical Depth.

		Period	**Slope (yr^{-1})**	**CrI or SD**
Borneo	AC(1) of spatially averaged VOD	Pre-breakpoint	0.008	CrI: 0.006-0.010
		Post-breakpoint	0.032	CrI: 0.029-0.034
	AC(1) of VOD for individual grid cells	Pre-breakpoint mean	0.00066	SD: 0.029
		Post-breakpoint mean	0.014	SD: 0.037
Amazon	AC(1) of spatially averaged VOD	Post-breakpoint	0.014	CrI: 0.012-0.016

(b) Date of breakpoint for spatially-averaged VOD AC(1) time series of Borneo and the Amazon. CrI: Credible Interval.

	Breakpoint Date	**CrI**
Borneo	Mar. 2000	Dec. 1999 – Jul. 2000
Amazon	Jan. 2000	Nov. 1999 – Mar. 2000

Table D.2: Spearman's rank correlation coefficients (ρ) and associated p-values between resilience indicators, metrics and periods. Unless otherwise specified, Kendall's τ and slope values are for VOD AC(1) time series. Abbreviations: AC(1) - lag-1 autocorrelation; VOD - Vegetation Optical Depth.

(a) Various correlations for Borneo.

	Correlation between	ρ	***p*-value**
1	AC(1) and variance of spatially-averaged VOD, full period	0.459	<0.001
2	AC(1) and variance of spatially-averaged VOD, post-breakpoint	0.533	<0.001
3	Individual grid cells' Kendall's τ and slope, pre-breakpoint	0.923	<0.001
4	Individual grid cells' Kendall's τ and slope, post-breakpoint	0.914	<0.001
5	Breakpoint timing and individual grid cells' pre-breakpoint Kendall's τ	0.145	0.002
6	Breakpoint timing and individual grid cells' pre-breakpoint slope	0.135	0.003
7	Breakpoint timing and individual grid cells' post-breakpoint Kendall's τ	0.084	0.007
8	Breakpoint timing and individual grid cells' post-breakpoint slope	0.15	0.001
9	Individual grid cells' Kendall's τ, pre- and post-breakpoint	-0.587	<0.001
10	Individual grid cells' slope, pre- and post-breakpoint	-0.586	<0.001

(b) Correlations between pre- and post-breakpoint VOD AC(1) consistency (Kendall's τ) and magnitude (slope) for grid cells of the Amazon rainforest.

Correlation between	ρ	***p*-value**
Individual grid cells' Kendall's τ, pre- and post-breakpoint	-0.603	<0.001
Individual grid cells' slope, pre- and post-breakpoint	-0.532	<0.001

Table D.3: Spearman's rank correlation coefficients (ρ) and associated p-values between pre- and post-breakpoint values of Kendall's τ and slope for the spatially-averaged VOD AC(1) of sliding bands of grid cells advancing along gradients of various factors (i.e., between blue and red dots in Figs. 5.3, D.9, D.10, D.11).

(a) Correlations for main factors using sliding bands of 100 grid cells.

Factors	**Metrics**	ρ	p-value
Precipitation	Kendall's τ	-0.749	<0.001
	Slope	-0.550	<0.001
Temperature	Kendall's τ	-0.717	<0.001
	Slope	-0.683	<0.001
Human Footprint	Kendall's τ	-0.496	<0.001
	Slope	-0.151	0.003
Biodiversity	Kendall's τ	0.250	<0.001
	Slope	0.060	0.237

(b) Correlations for individual taxa using sliding bands of 100 grid cells.

Factors	**Metrics**	ρ	p-value
Amphibians	Kendall's τ	0.269	<0.001
	Slope	0.446	<0.001
Birds	Kendall's τ	-0.716	<0.001
	Slope	-0.556	<0.001
Mammals	Kendall's τ	-0.820	<0.001
	Slope	-0.434	<0.001
Reptiles	Kendall's τ	-0.241	<0.001
	Slope	-0.081	0.111

(c) Correlations for main factors using sliding bands of 50 grid cells.

Factors	**Metrics**	ρ	p-value
Precipitation	Kendall's τ	-0.577	<0.001
	Slope	-0.341	<0.001
Temperature	Kendall's τ	-0.583	<0.001
	Slope	-0.586	<0.001
Human Footprint	Kendall's τ	-0.518	<0.001
	Slope	-0.084	0.081
Biodiversity	Kendall's τ	-0.460	<0.001
	Slope	-0.263	<0.001

(d) Correlations for individual taxa using sliding bands of 50 grid cells.

Factors	**Metrics**	ρ	p-value
Amphibians	Kendall's τ	-0.124	0.010
	Slope	-0.179	<0.001
Birds	Kendall's τ	-0.606	<0.001
	Slope	-0.363	<0.001
Mammals	Kendall's τ	-0.587	<0.001
	Slope	-0.451	<0.001
Reptiles	Kendall's τ	-0.346	<0.001
	Slope	-0.170	<0.001

Table D.4: Results of robustness tests for early warning signals, lag-1 autocorrelation (AC(1)) and variance, calculated on the acoustic complexity and bioacoustic indices from recordings along a land-use gradient in Malaysian Borneo. Abbreviations: Av - Average; Var - Variance.

(a) Acoustic indices calculated on 10 min periods, with a 2 hr window advancing in 30 min increments.

		Old-Growth	Logged	Oil Palm
Acoustic Complexity	**Av. AC(1)**	0.330	0.355	0.510
	Av. Var.	52.3	68.3	13.3
Bioacoustic	**Av. AC(1)**	0.641	0.439	0.509
	Av. Var.	3411	575	1113

(b) Acoustic indices calculated on 10 min periods, with a 6 hr window advancing in 2 hr increments.

		Old-Growth	Logged	Oil Palm
Acoustic Complexity	**Av. AC(1)**	0.510	0.570	0.660
	Av. Var.	43.6	73.5	16.2
Bioacoustic	**Av. AC(1)**	0.859	0.549	0.629
	Av. Var.	6158	493	1320

(c) Acoustic indices calculated on 10 min periods for full-length average time series at each site, with a 3 hr window advancing in 1 hr increments (as in main text).

		Old-Growth	Logged	Oil Palm
Acoustic Complexity	**Av. AC(1)**	0.407	0.454	0.58
	Av. Var.	53.5	55.8	10
Bioacoustic	**Av. AC(1)**	0.714	0.448	0.533
	Av. Var.	4640	774	882

(d) Acoustic indices calculated on 1 min periods, with a 3 hr window advancing in 1 hr increments (as in main text).

		Old-Growth	Logged	Oil Palm
Acoustic Complexity	**Av. AC(1)**	0.265	0.136	0.177
	Av. Var.	53.2	117	13.6
Bioacoustic	**Av. AC(1)**	0.596	0.475	0.362
	Av. Var.	4689	812	1468

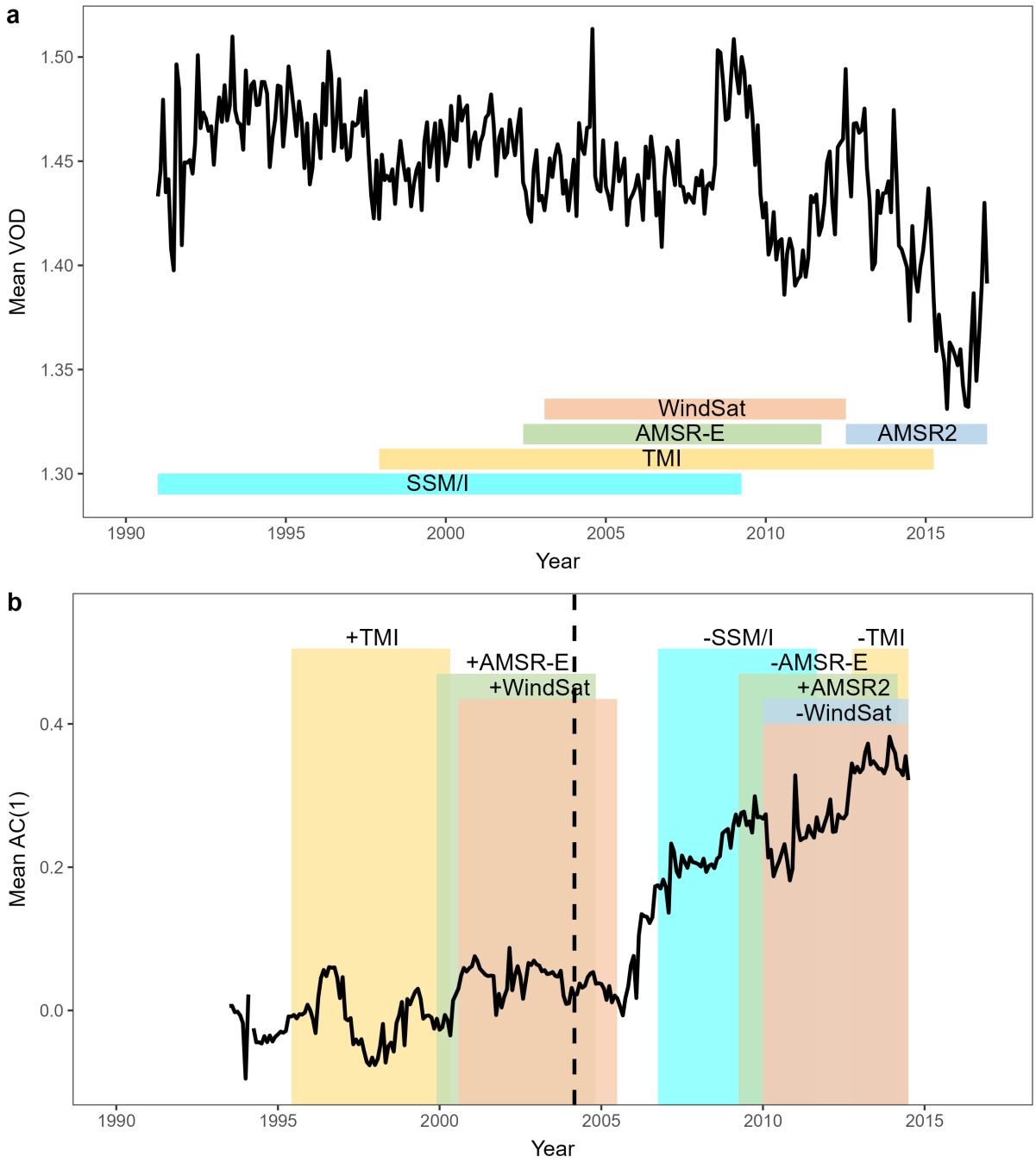


Figure D.1: Relative time and impact of sensor shifts on time series of VOD and AC(1). (**a**), Time series of monthly mean VOD spatially averaged over the study area, with the time span of each satellite sensor used to measure VOD shown in the bottom, adapted from Moesinger et al. (2020). (**b**), Mean AC(1) time series created from mean VOD time series above, calculated on 5-year sliding windows with monthly increments and plotted at the midpoint of each window. The coloured bands correspond to the addition (+) or removal (−) of sensors and highlight sections of the AC(1) time series that were calculated on windows of VOD data that feature those particular sensor changes. The vertical dotted line indicates the breakpoint of AC(1) time series.

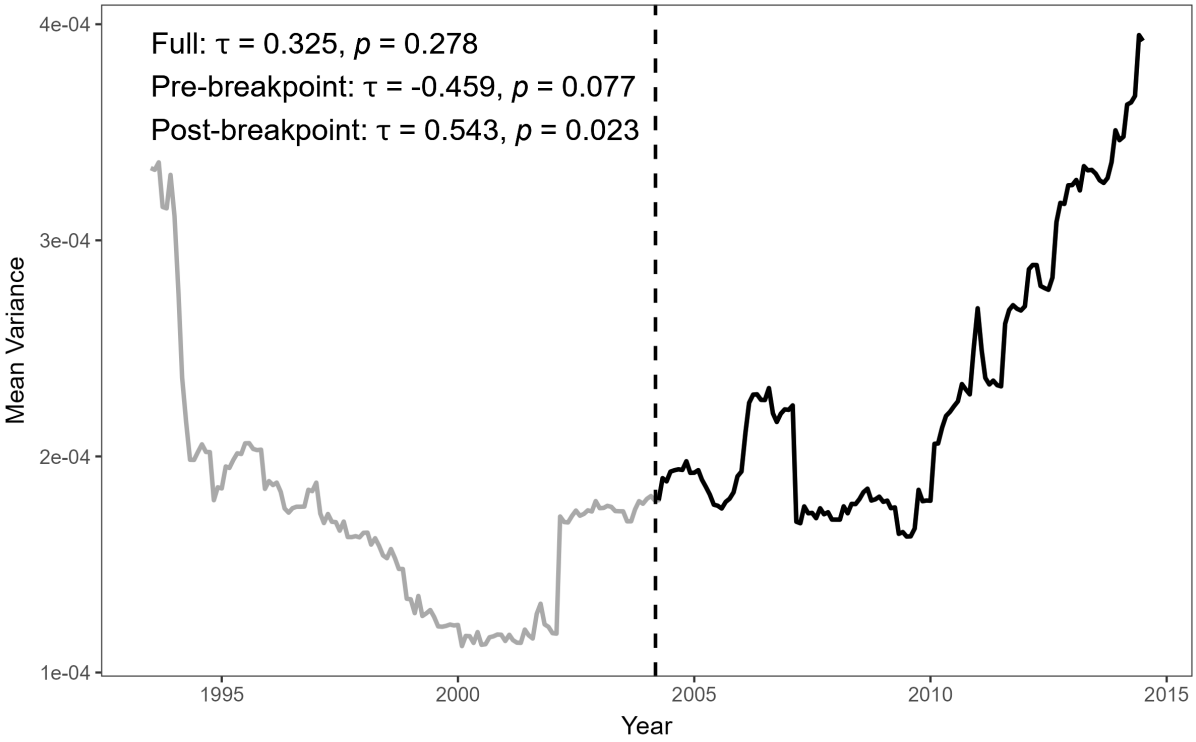


Figure D.2: Changes in satellite-derived Bornean rainforest vegetation resilience since the 1990s, as measured from the variance of VOD. The plotted time series corresponds to the mean VOD variance for the study area from July 1993 to July 2014. The vertical dotted line indicates the breakpoint determined from the VOD AC(1) time series (approximately March 2004; see Fig. 5.1). The Kendall's τ values and their p-values for the full study period and pre- and post-breakpoint periods (i.e., before and after the dotted line) are noted in the upper left quadrant of the figure, indicating the consistency of resilience change for the different periods (negative values suggest resilience gain, positive values suggest resilience loss). Note that the variance values are plotted at the middle of each 5-year sliding window.

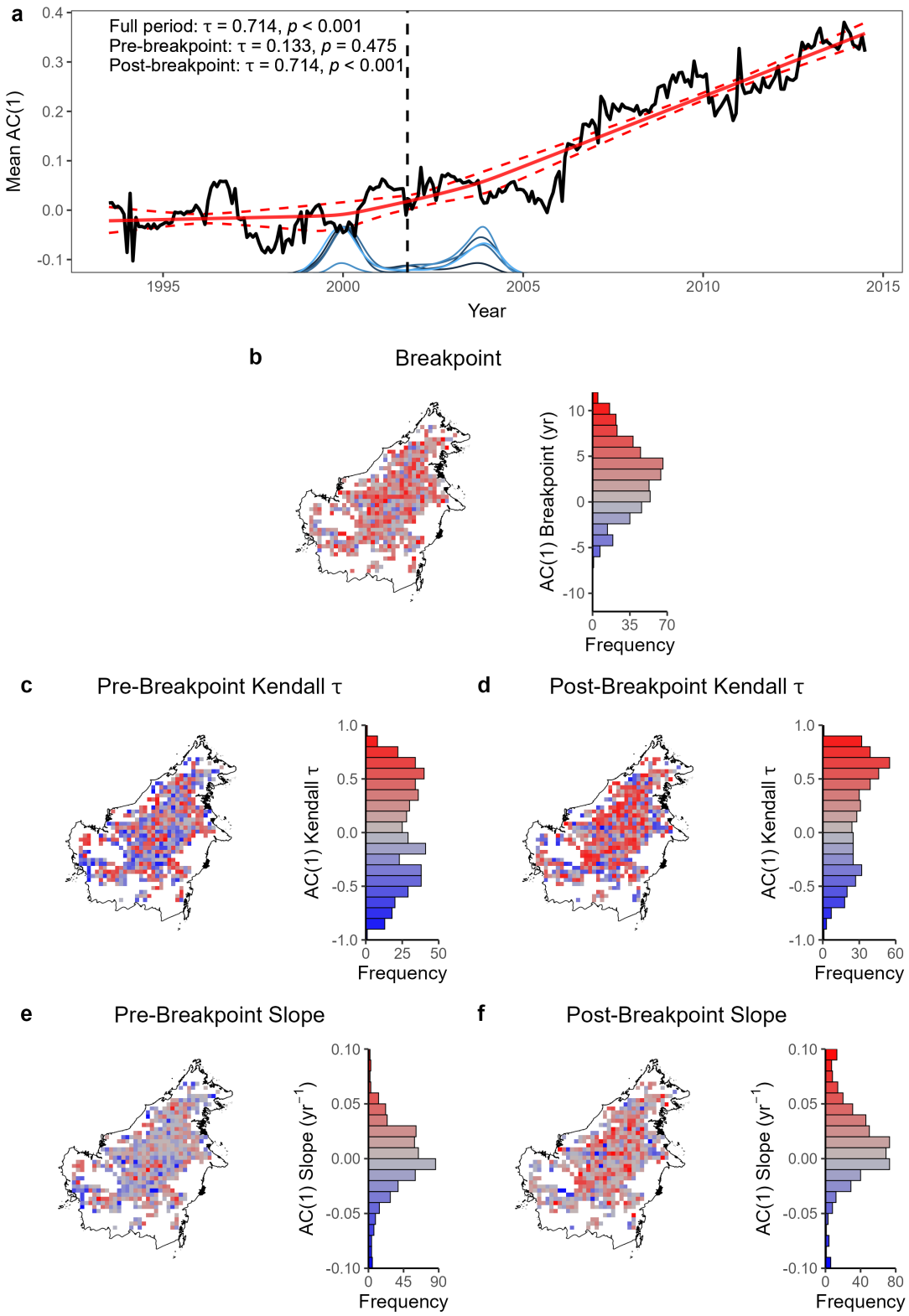


Figure D.3: Overall trend and spatial patterns in resilience changes of Bornean rainforest when varying the threshold of evergreen broadleaf forest cover to 60% to determine the study area (i.e., recreation of Fig. 5.1a and Fig. 5.2b-f but using grid cells with forest cover $\geq$ 60%).

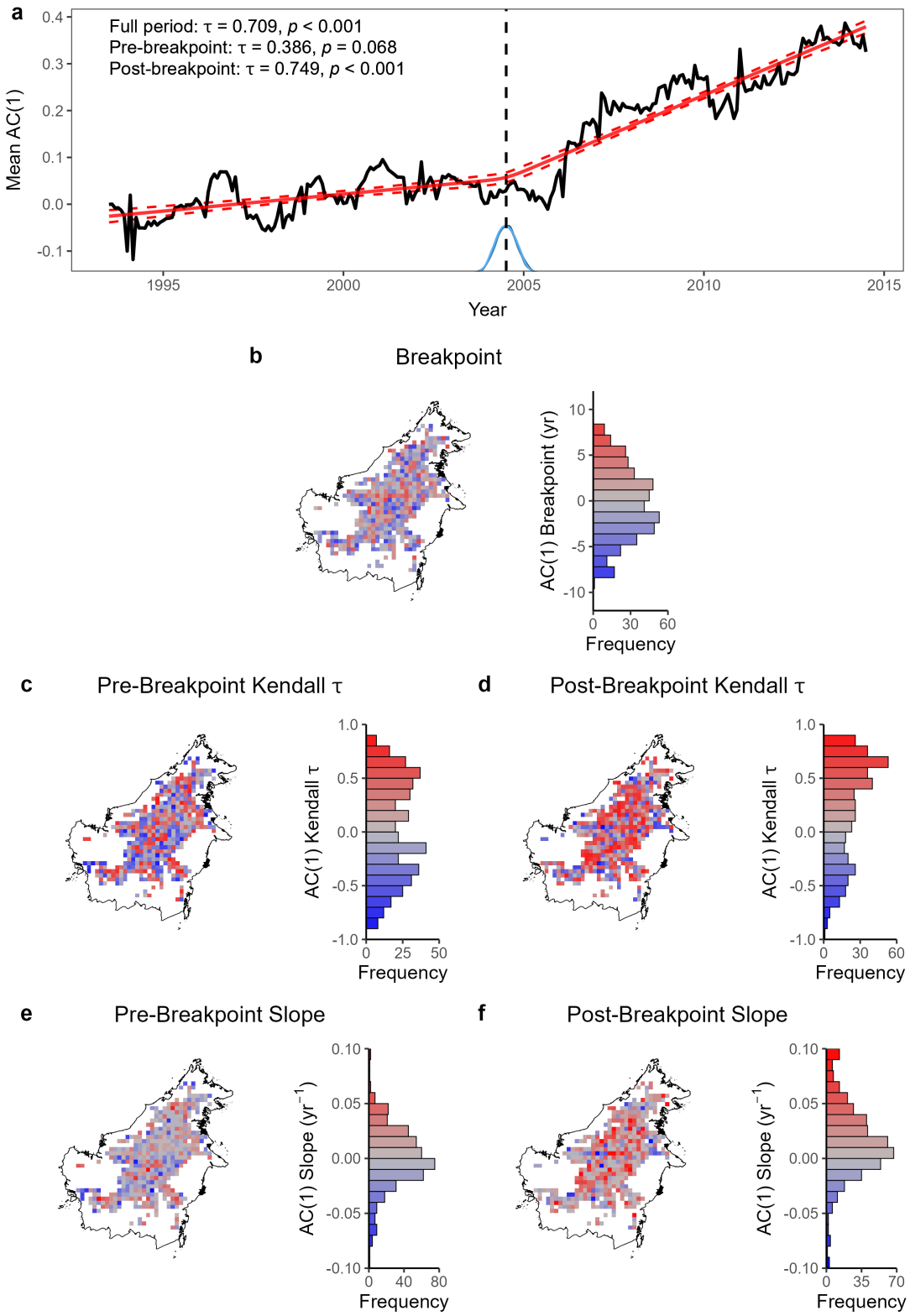


Figure D.4: Overall trend and spatial patterns in resilience changes of Bornean rainforest when varying the threshold of evergreen broadleaf forest cover to 90% to determine the study area (i.e., recreation of Fig. 5.1a and Fig. 5.2b-f but using grid cells with forest cover $\geq$ 90%).

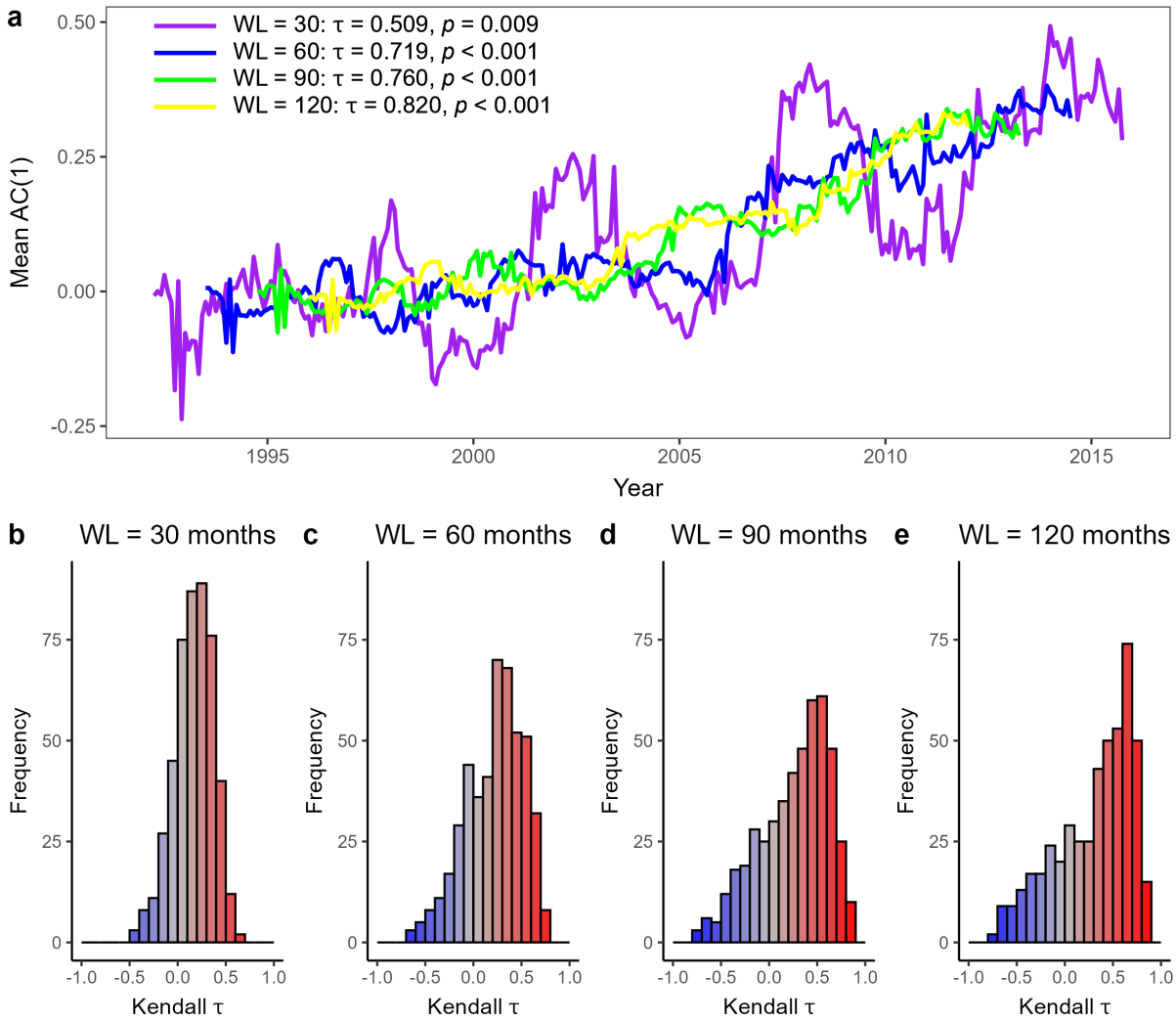


Figure D.5: Robustness of VOD AC(1) to VOD window length. (**a**), Mean AC(1) time series when varying the VOD window length to 30 months, 60 months (as in Fig. 5.2d), 90 months and 120 months. Note that the AC(1) values are plotted in the middle of the sliding window. (**b**–**e**), Histograms of grid cells' Kendall's τ values for these choices of AC(1) window lengths. WL - window length.

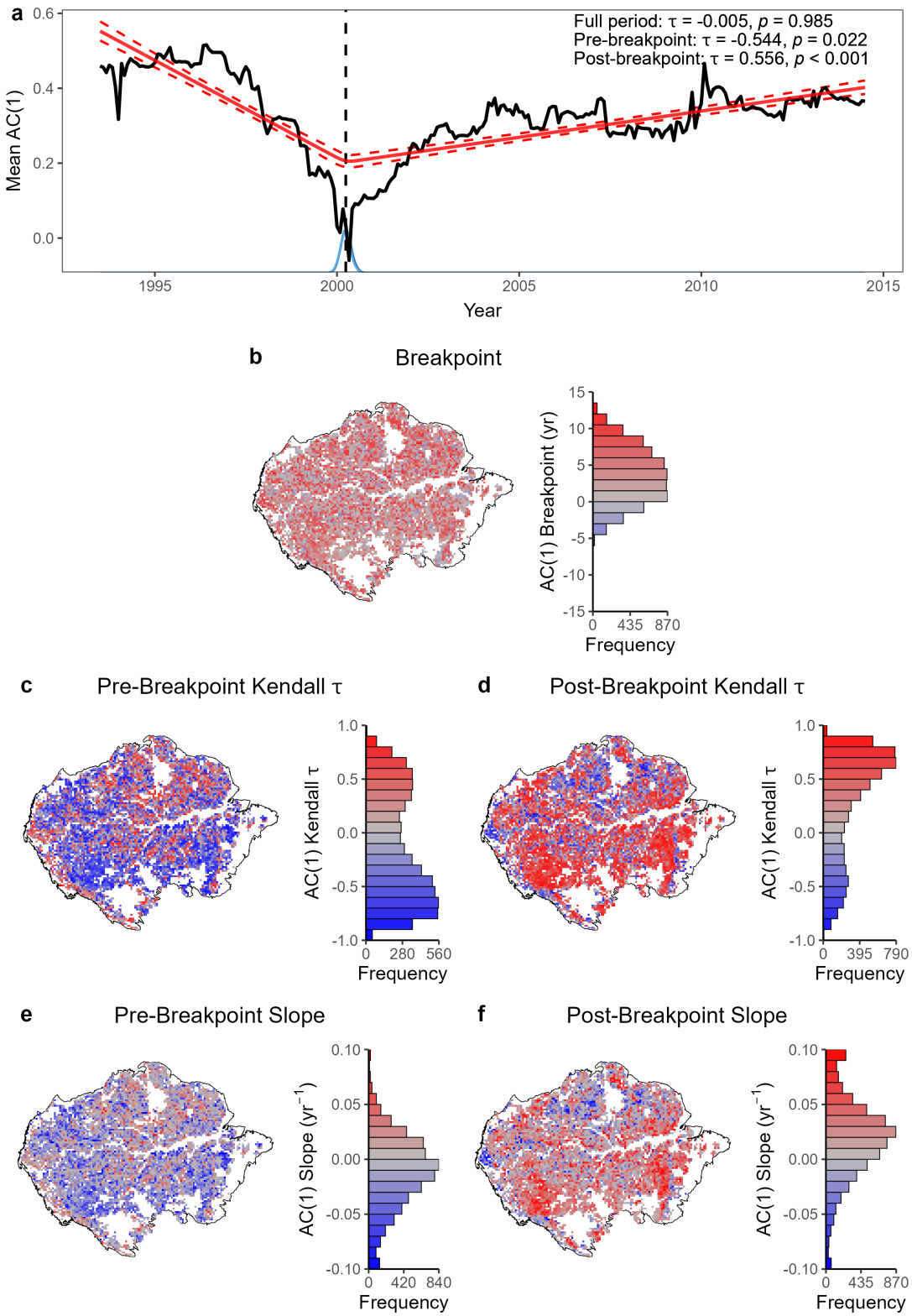


Figure D.6: Overall trend and spatial patterns in resilience changes of Amazon rainforest vegetation, adapted from Boulton et al. (2022) with our processing pipeline. Recreation of Fig. 5.1a and Figs. 5.2b–f but for the study area of the Amazon basin. Note that in our processing pipeline, VOD data are spatially averaged before calculating the overall mean AC(1) time series plotted in panel (**a**).

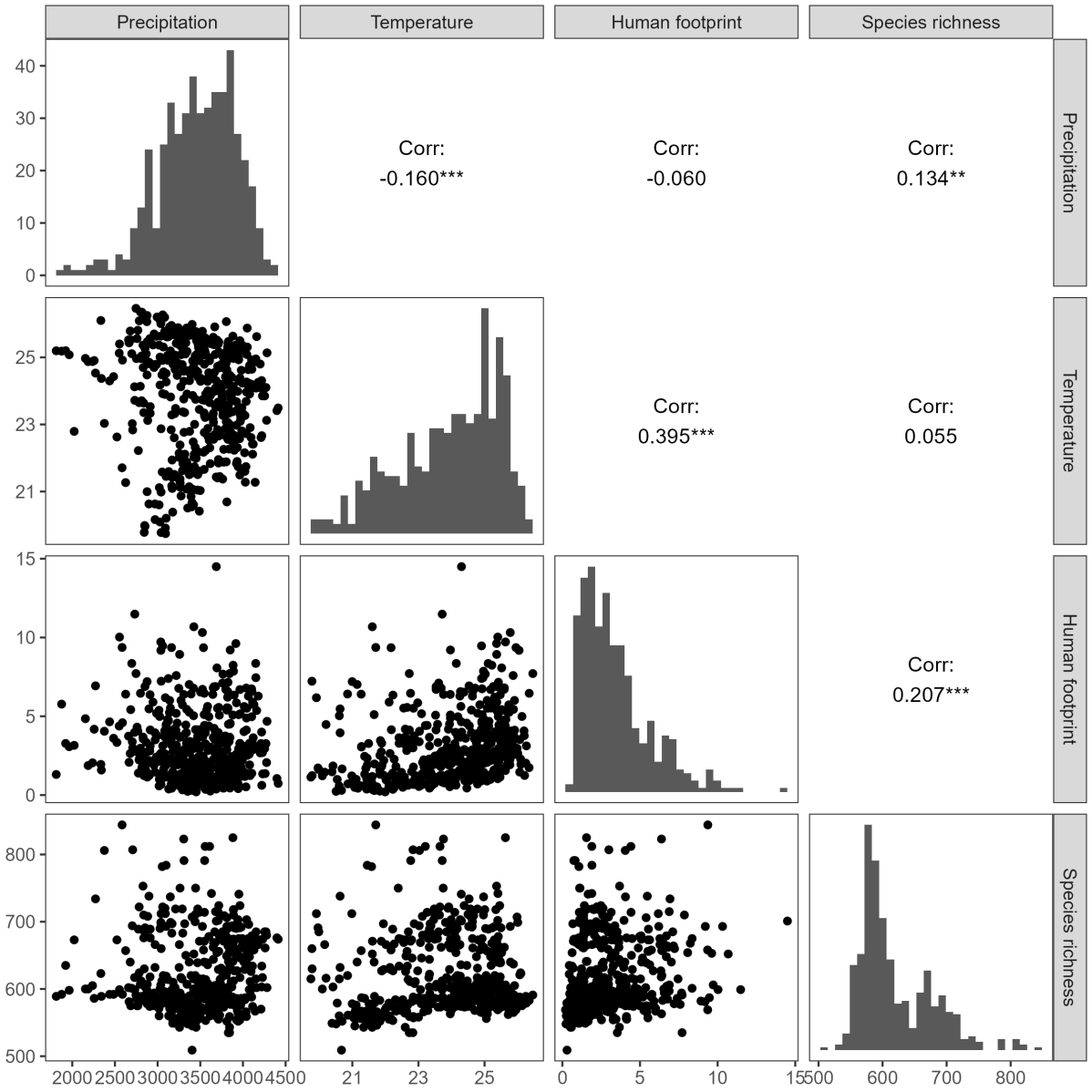


Figure D.7: Pairwise correlations between environmental factors. The upper right panels display Spearman's correlation coefficients with significance levels (***: $p \leq 0.001$, **: $p \leq 0.01$, *: $p \leq 0.05$). The diagonal panels show histograms of each factor's distribution, while the bottom left panels present scatter plots visualizing pairwise trends.

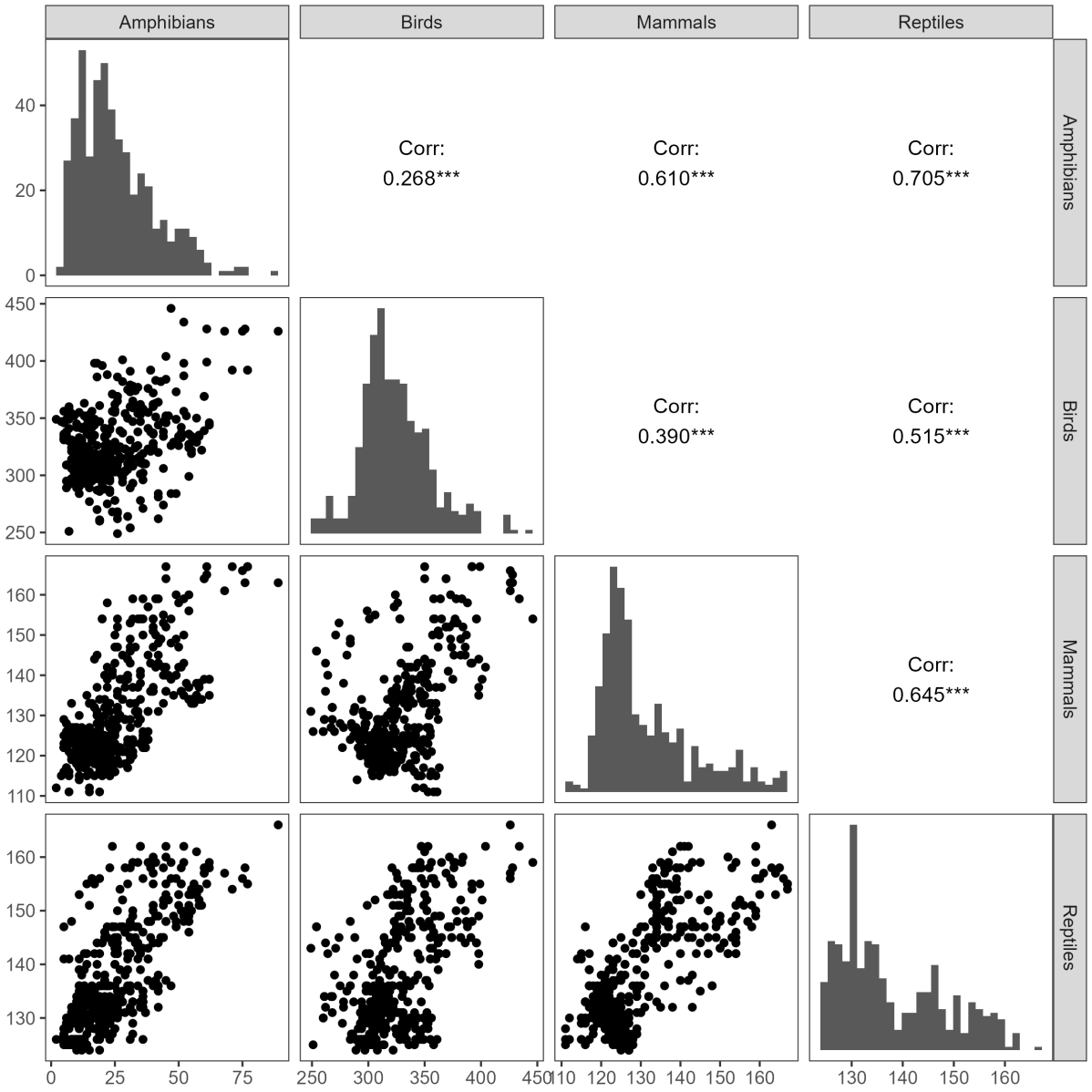


Figure D.8: Pairwise correlations between taxa. The upper right panels display Spearman's correlation coefficients with significance levels (***: $p \leq 0.001$, **: $p \leq 0.01$, *: $p \leq 0.05$). The diagonal panels show histograms of each taxon's distribution, while the bottom left panels present scatter plots visualizing pairwise trends.

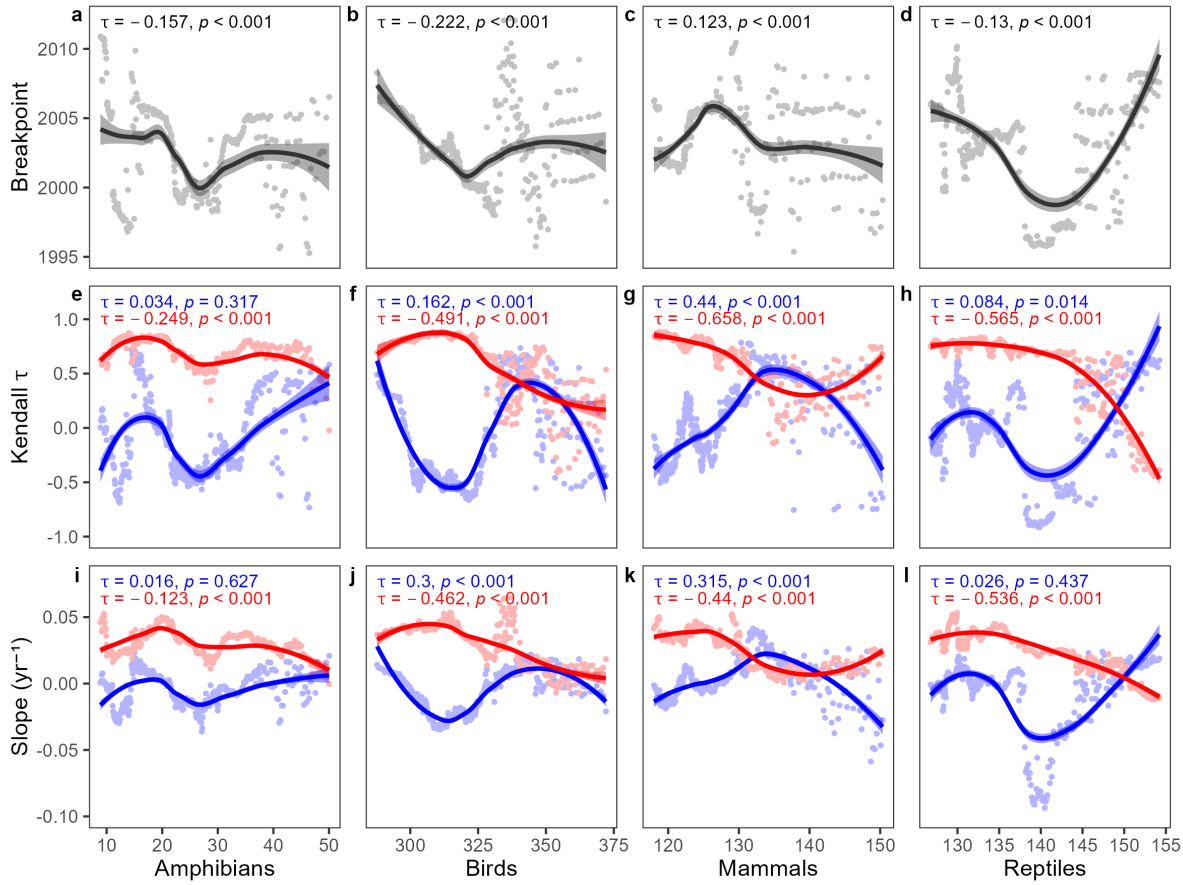


Figure D.9: Relationship between vegetation resilience and species richness of different animal taxa (sliding band size = 100 cells). Rows correspond to different resilience metrics (breakpoint – timing of accelerated resilience loss; Kendall's τ – consistency of resilience change; slope – magnitude of resilience change; Section 5.2.1) and columns represent different factors. (**a, e, i**) Correlation between amphibian richness and the following resilience metrics calculated on sliding bands of grid cells along a species richness gradient of amphibians: (**a**) posterior mean breakpoint from Bayesian breakpoint regressions; (**e**) Kendall's τ (blue: pre-breakpoint; red: post-breakpoint); and (**i**) posterior mean slope (blue: pre-breakpoint; red: post-breakpoint). The correlations between the resilience metrics and potential influencing factors are estimated using Kendall's τ, which is reported with its associated p-values at top of each panel. (**b–d, f–h, j–l**) The same as (**a, e, i**), but for (**b, f, j**) birds, (**c, g, k**) mammals and (**d, h, l**) reptiles, respectively. Species richness data are from the IUCN Red List [317] (Section 5.2.1). In each panel, the values are plotted at the midpoint of each band.

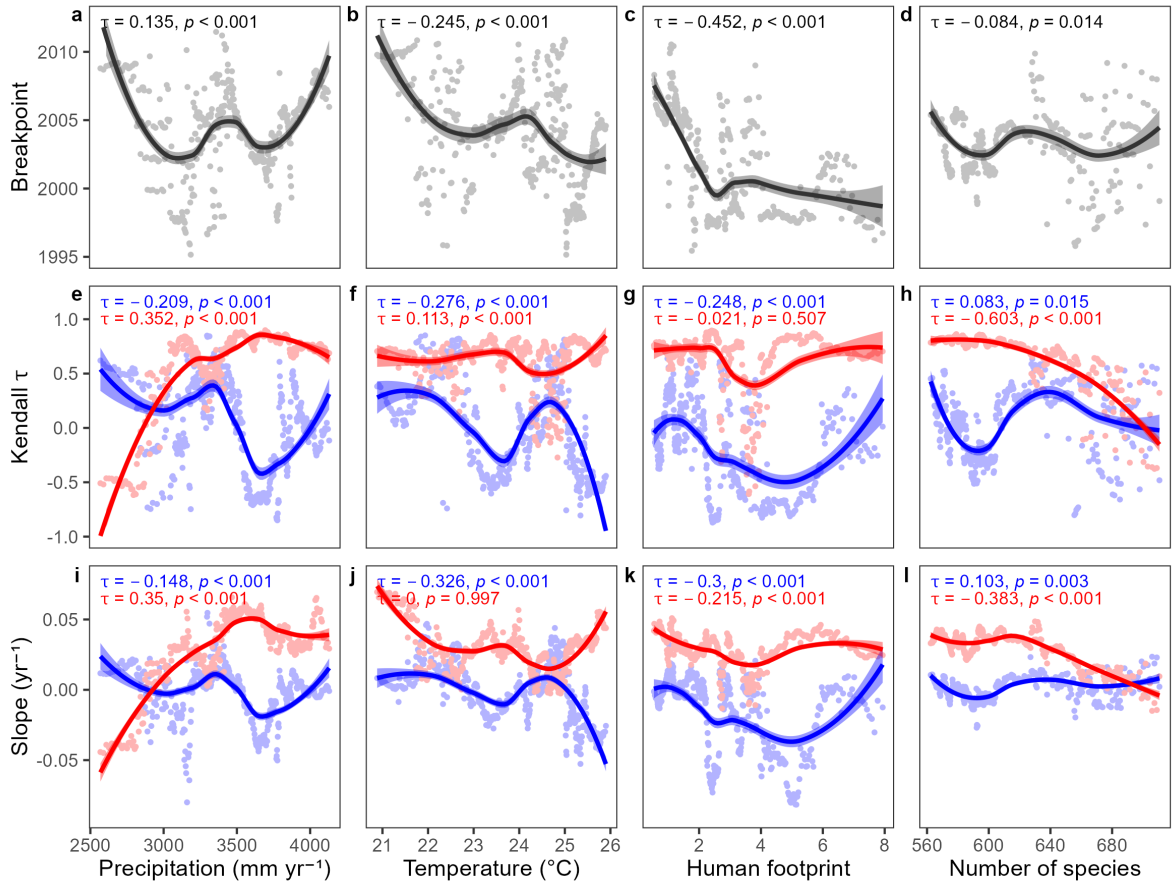


Figure D.10: Relationship between potential influencing factors and Bornean rainforest vegetation resilience when varying the band size to 50 cells (recreation of Fig. 5.3).

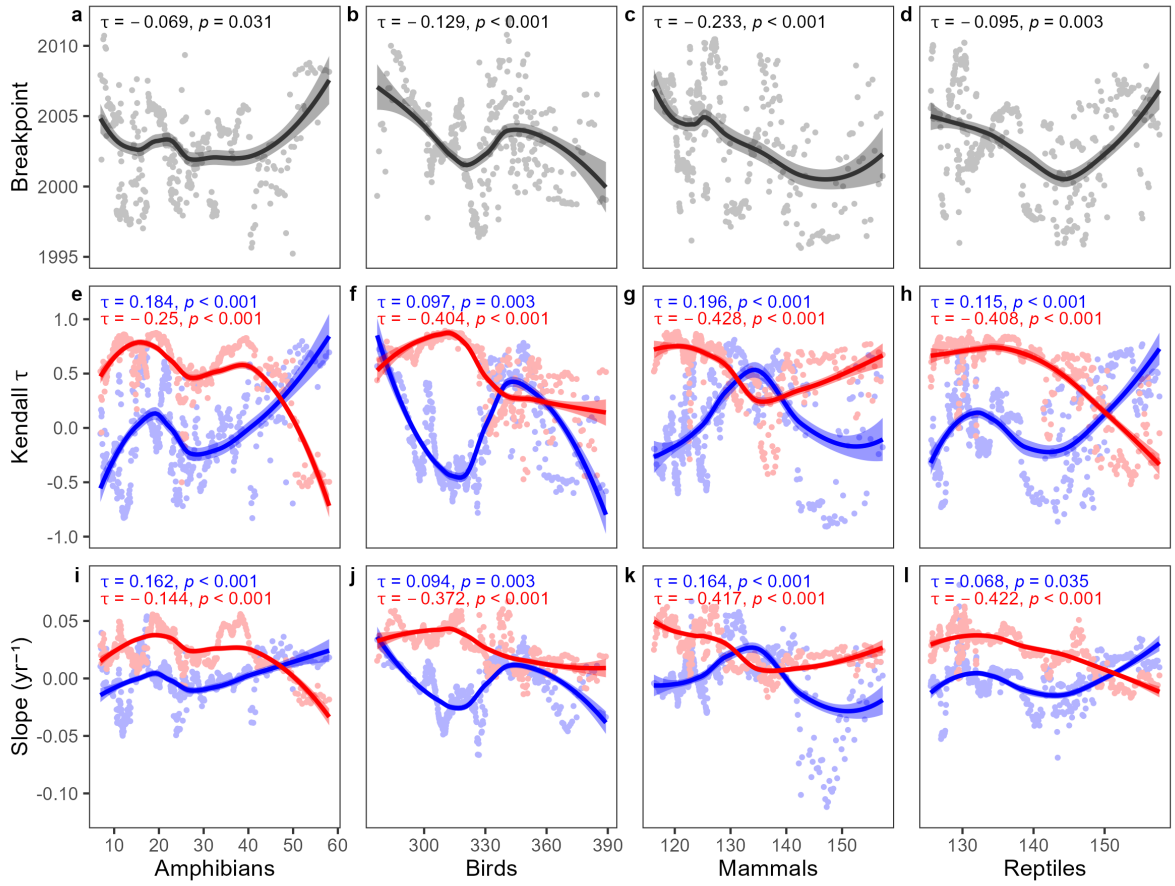


Figure D.11: Relationship between species richness of different animal taxa and resilience when varying the band size to 50 cells (recreation of Fig. D.8).

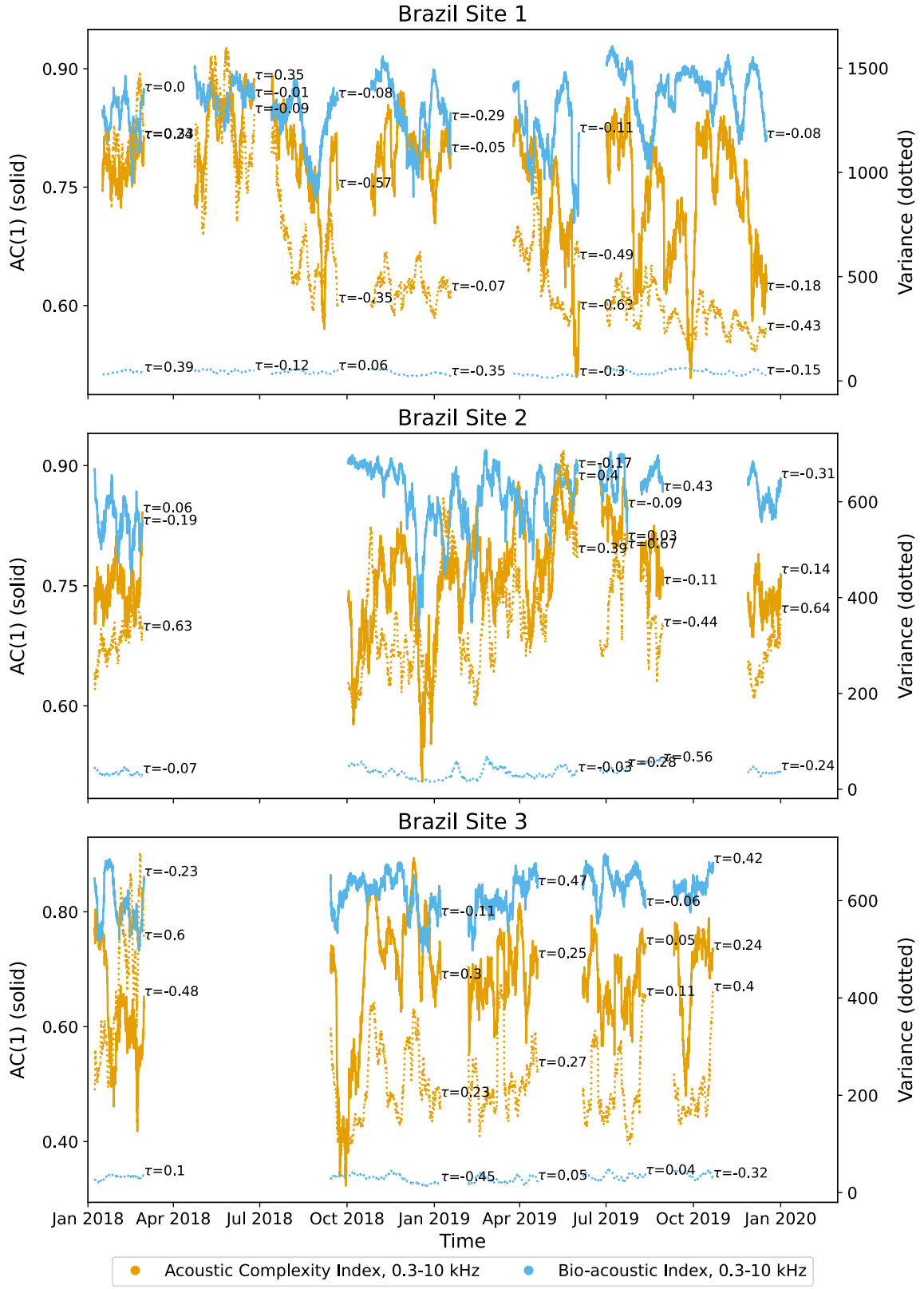


Figure D.12: Recreation of Fig. 5.5, calculating early warning signals on one-week-long windows of acoustic indices advancing hourly.

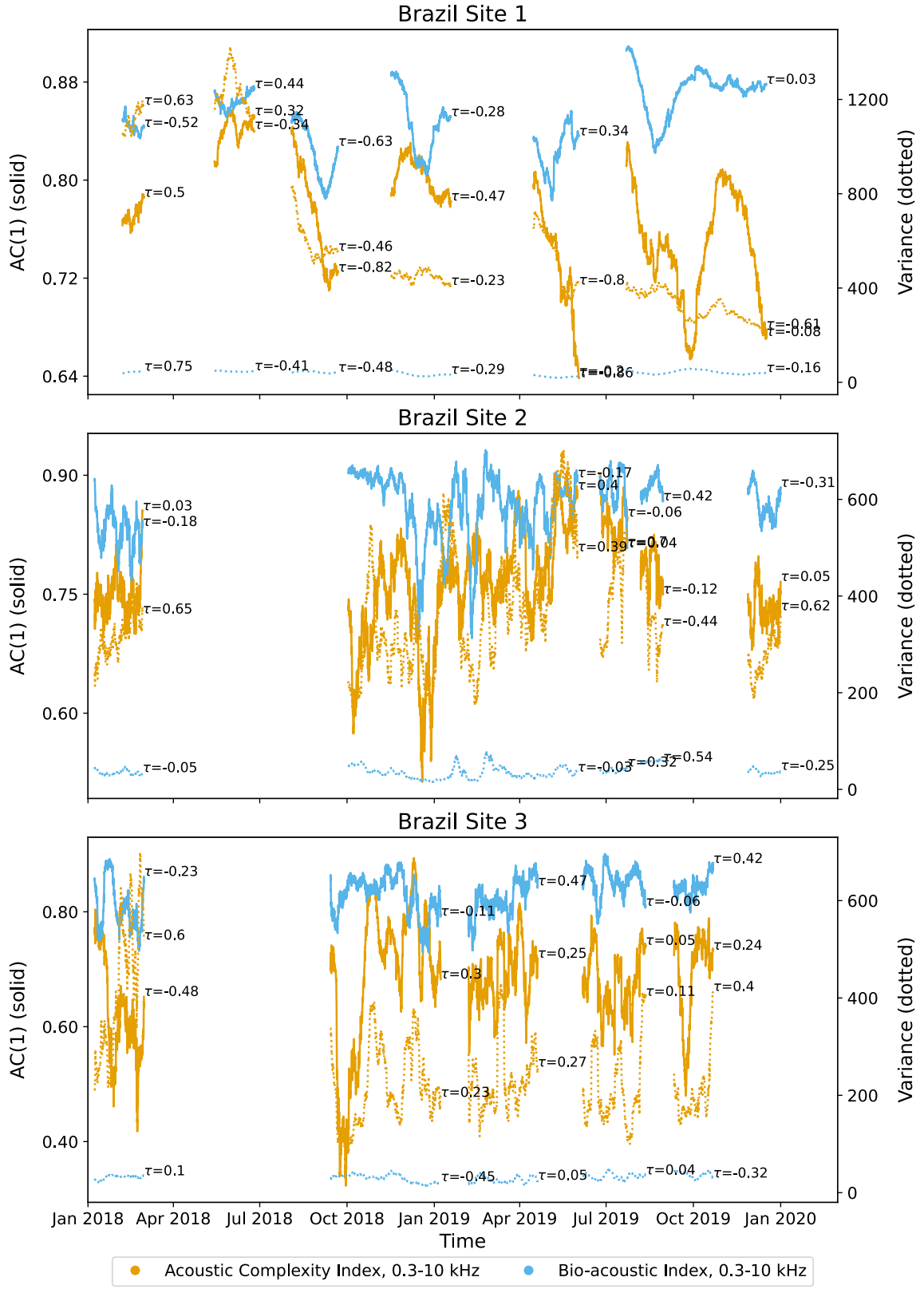


Figure D.13: Recreation of Fig. 5.5, calculating early warning signals on four-week-long windows of acoustic indices advancing hourly.

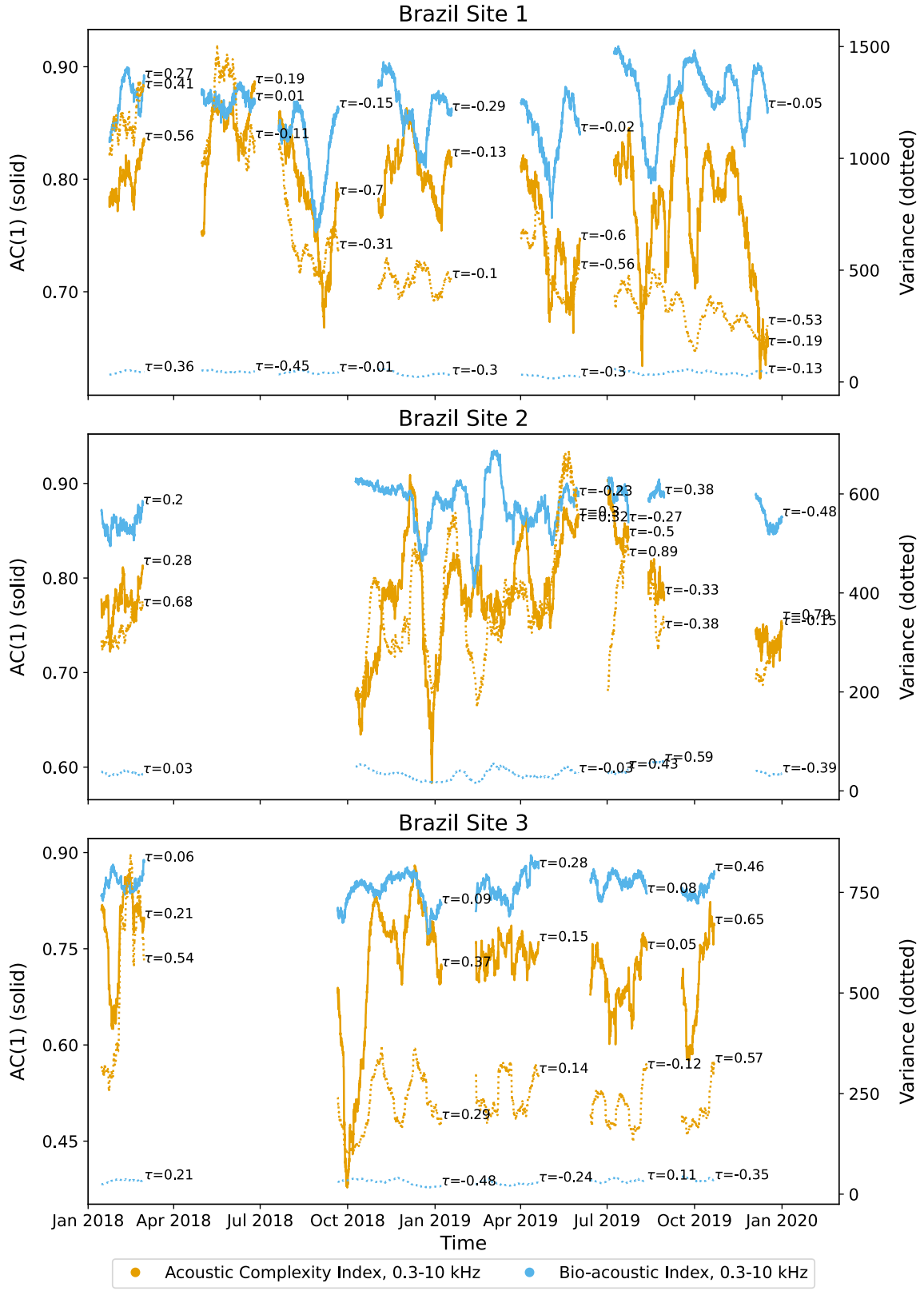


Figure D.14: Recreation of Fig. 5.5, calculating early warning signals on acoustic indices for which data gaps were filled by linear interpolation.

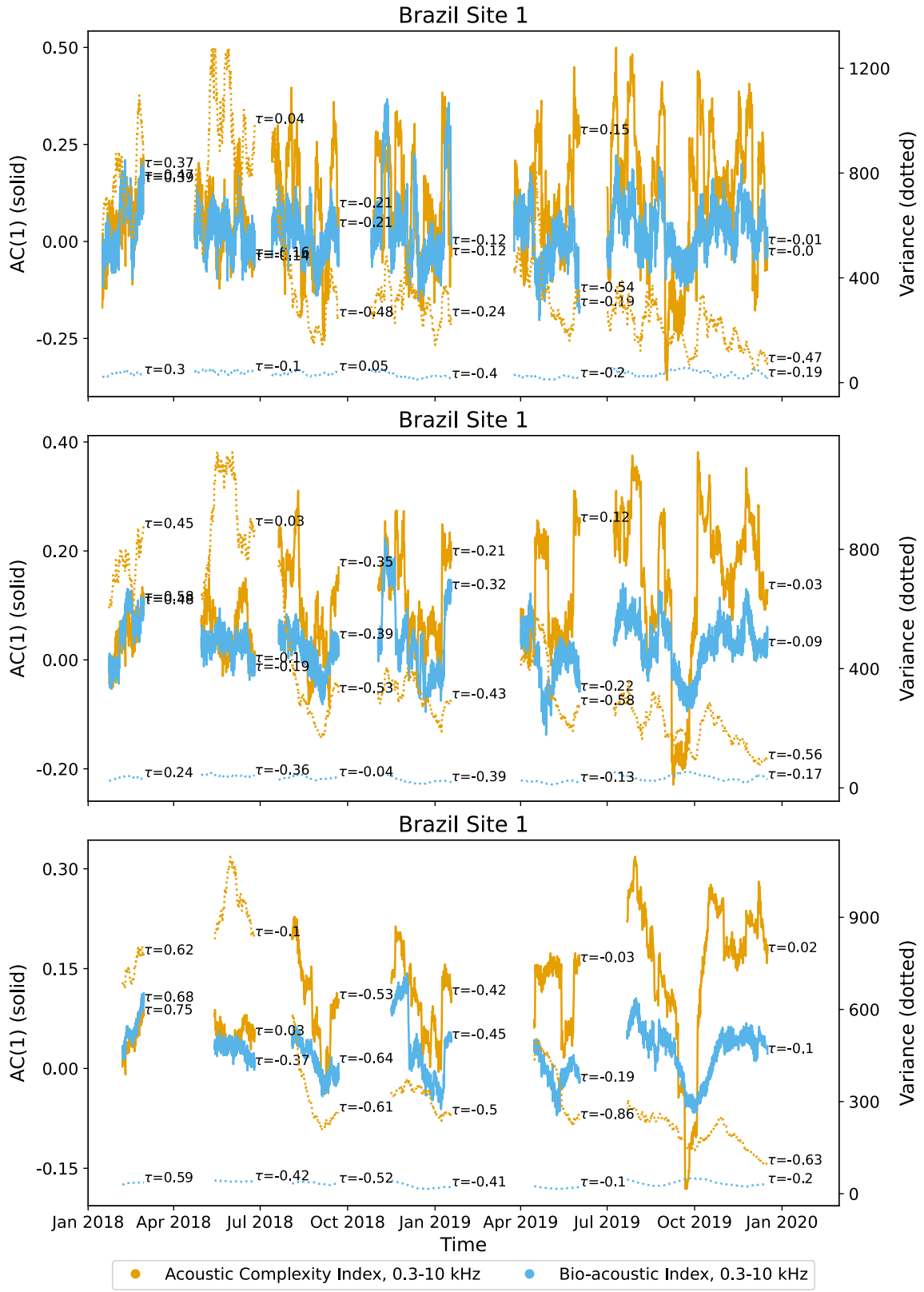


Figure D.15: Recreation of Fig. 5.5 Site 1, calculating early warning signals on acoustic indices resampled to 6 hr resolution using windows that are 1 week (top panel), 2 weeks (middle panel) and 4 weeks (bottom panel) long and advance in 6 hr increments.