A Proof-of-Concept for Acoustic Monitoring of Breeding Phenology and Success in Cavity-Nesting Migrants

Andrew John Heppenstall Doyle

How to cite:

Doyle, Andrew John Heppenstall (2026) A Proof-of-Concept for Acoustic Monitoring of Breeding Phenology and Success in Cavity-Nesting Migrants. Masters thesis, Durham University.

A Proof-of-Concept for Acoustic Monitoring of Breeding Phenology and Success in Cavity-Nesting Migrants

Andrew John Heppenstall Doyle

Submitted for the Degree of Masters by Research

Supervised by Dr Christine Howard and Prof Steve Willis

Department of Biosciences

Durham University

2025

Abstract

Understanding how migratory bird populations respond to environmental change requires linking migratory behaviour to reproductive outcomes. For long-distance migrants such as the Pied Flycatcher (*Ficedula hypoleuca*), timely arrival at breeding grounds is critical for reproductive success, yet models predicting population dynamics lack sufficient empirical data on how breeding phenology and success vary between migration strategies.

This thesis presents a proof-of-concept for using passive acoustic monitoring to detect breeding behaviour and estimate breeding phenology and success metrics in cavity-nesting migrants. Traditional methods, including colour ringing, tracking devices, and in-person nest box monitoring, are limited by costs and labour demands. Audio recorders offer a minimally invasive and scalable alternative, though their application to nest box monitoring remains largely untested.

To address this, two complementary studies were conducted. First, the feasibility of detecting nest box behaviours was evaluated using audio recorders and custom-built classifiers in Kaleidoscope Pro, validated using in-field observations. Song was detected with higher accuracy compared to calls and visitations which required greater manual validation. Comparisons with BirdNET indicated that custom classifiers captured more detections but required additional development and processing time. Second, Bayesian changepoint models were applied to behavioural detections over time to assess whether shifts in frequency aligned with in-field estimates for breeding stage transitions and success metrics. Vocalisation-derived changepoints reflected arrival timing, while visitation detections aligned with onset of chick-begging and fledging. Behaviour-derived estimates were consistent across full-day and reduced recording schedules, highlighting the potential for reductions in processing time and data storage. While visitation rates showed limited correlation with brood size, binary nest success or failure may be inferred.

These findings demonstrate that audio recorders provide reliable estimates of breeding phenology and represent a scalable complement to traditional monitoring. By connecting behaviour to reproductive phenology, acoustic monitoring offers novel empirical data to parameterise migration models, advancing predictive capacity under environmental change.

Table of Contents

Statement of Copyright

Acknowledgements

In completing this thesis, I would like to thank Dr Christine Howard and Professor Steve Willis for their supervision. I am also grateful to the Conservation Ecology Group at Durham University for their encouragement and support throughout the project.

For assistance with fieldwork, I would like to thank the Tees Ringing Group for allowing me to join their nest box monitoring project and for the invaluable guidance and knowledge they shared during data collection. I also extend my thanks to the landowners and managers for their support in granting access to the study site.

I am grateful to Wildlife Acoustics Inc. for awarding me their software grant, which enabled the processing of the acoustic data used in this project. I also acknowledge the financial support of the Durham University Postgraduate Student Support Scholarship, which provided me with the opportunity to focus on my research.

1. Thesis Introduction

1.1 Understanding the Threats to Long-distance Migrants

As seasons change throughout the year, available resources across the landscape change with it, causing birds worldwide to migrate in search of suitable habitats (Hahn et al., 2009). However, the once reliable seasonal distribution of resources that migrants depend on is changing due to increasing environmental uncertainty (Barbet-Massin et al., 2012; Both et al., 2006). Climate, land-use change, and overexploitation have all been linked with habitat loss, deterioration, or disturbance driving the population declines in migratory birds (Kubelka et al., 2022).

Population declines are particularly acute in long-distance migrants. Since 1980, trans-Saharan species such as the Pied Flycatcher (*Ficedula hypoleuca*) have declined steadily, in contrast to the short-distance migrants ability to recover from a population decline (Figure 1.1) (Berthold et al., 1998; Both et al., 2006; Goodenough et al., 2009; Howard et al., 2020). Consistent decline for *F. hypoleuca* and other Afro-Palearctic migrants breeding in the UK are largely linked to changes in their non-breeding grounds and along migratory routes (Figure 1.2) (Howard et al., 2020; Nater et al., 2023), where land cover suitability has decreased since 1980 despite relative stability in suitable land cover on the breeding grounds (Howard et al., 2020). Populations of long-distance migrants are particularly vulnerable as they are exposed to multiple threats to survival and reproductive fitness during migration. In contrast, short-distance migrants have closer breeding and non-breeding ranges, making their population dynamics more strongly driven by environmental conditions at those sites rather than by risks encountered en route (Howard et al., 2020).

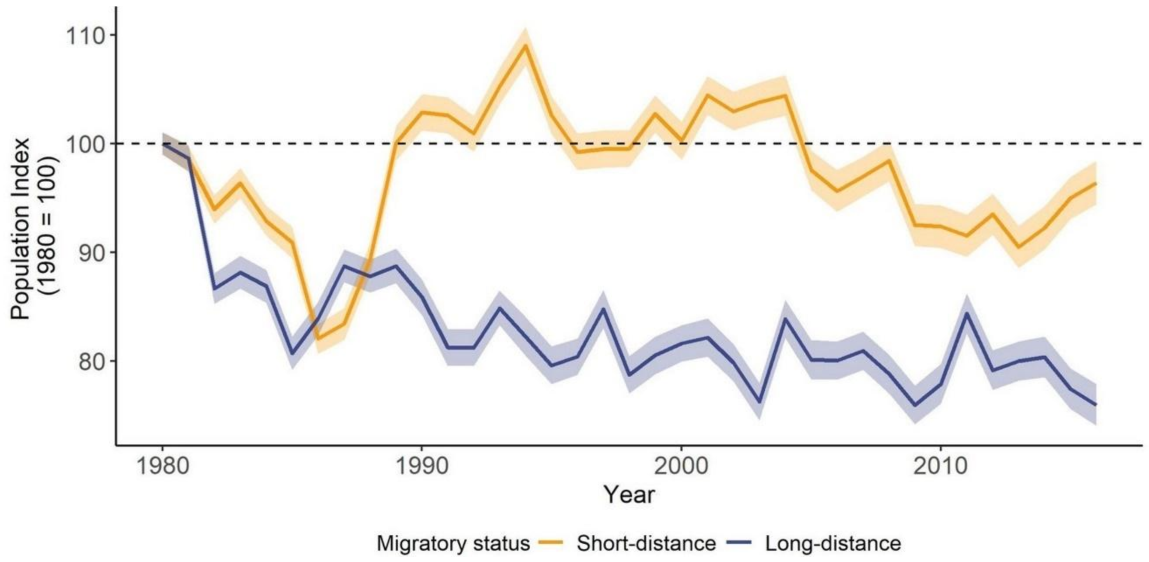

Figure 1.1 Changes in population indices of 61 short-distance (yellow line) and 39 long-distance (blue line) migrants from 1980 to 2016. Short-distance migrants, were defined as birds that migrate principally to different parts of Europe and North Africa in the non-breeding season; and long-distance migrants, birds that spend the non-breeding season entirely in sub-Saharan Africa. Solid lines depict the geometric mean, with shading representing the standard deviation. Population indices for 1980 are set to 100 for reference. The horizontal dashed line at the value of 100 indicate the expectation if there is no change. Figure taken from Howard et al., 2020.

Migration is a complex and risky process, exposing individuals to threats such as predation, adverse weather, and resource scarcity – pressures that are intensifying under environmental change (Buchan et al., 2022; Ozsanlav-Harris et al., 2023). Strategies that vary among populations and individuals can buffer against localised or stochastic events, but this variation is increasingly constrained as conditions change (Adamík et al., 2023; Bell et al., 2022; Howard et al., 2020; Ouwehand et al., 2016). Climate change is predicted to reduce habitat availability at wintering grounds and along migratory routes. Poleward shifts in breeding and non-breeding habitats will increase in migratory journeys and their reliance on diminishing stopover sites (Barbet-Massin et al., 2009; Howard et al., 2018, 2020). As a result, long-distance migrants may become more dependent on fewer stopover sites at a time in the annual cycle when interspecific competition already limits resource access (Zimin et al., 2023).

Migratory decisions are also constrained by the phenological timing of resource availability along the migratory route and at the breeding grounds. Climate and land-use change are increasing uncertainty in resource availability and advancing seasonal emergence of resources, with consequences for migratory behaviour, refuelling, and survival (Shepard et al., 2013). Limited food availability at wintering grounds can delay departure and hinder adaptation to earlier migration schedules (Helm et al., 2019; Ouwehand et al., 2023), while deteriorating or unpredictable conditions en route may prolong stopovers and elevate starvation risk (Arlt et al., 2015; Cowley & Siriwardena, 2005; Tomotani et al., 2018).

Delays during migration compromise arrival timing and an individual's condition at the breeding ground, with negative carry-over effects on breeding performance (Howard et al., 2018; Newton, 2006; Zurell et al., 2018). Increasingly, arrival is mismatched with the phenological advance of breeding ground resources, reducing access to optimal territories as late migrants face competition from resident species (Goodenough, 2014; Samplonius et al., 2018; Vriend et al., 2023; Wittwer et al., 2015). As fledging success depends on territory quality, delayed arrival poses significant risks for reproductive success (Nicolau et al., 2021; Shutt et al., 2022), and this effect may be reinforced across generations, as hatching is linked to their spring departure as an adult (Wang et al., 2022).

For long-distance migrants like *F. hypoleuca*, the ability to adjust migration strategies is critical for arriving on time and in good condition for reproduction (Both et al., 2006; Ozsanlav-Harris et al., 2023). However, the extent to which individuals can adapt their migratory strategies is uncertain, with potential consequences for breeding success and population viability.

Understanding how *F. hypoleuca* population dynamics respond to increasing environmental uncertainty throughout their annual cycle requires linking migratory decisions, such as route and timing, to reproductive outcomes. Novel migration models capable of predicting migratory journeys and population dynamics under environmental change, such as the Spatially-explicit Adaptive

Migration Model (SAMM) incorporate this link through the terminal reward function – a parameter which defines how an individual's state at the end of migration relates to their reproductive fitness (Clark & Mangel, 2000; Houston & McNamara, 1999; Howard et al., 2024). Model outputs are therefore sensitive to how the terminal reward is defined, accurate simulations depend on empirical data to describe how timing of arrival influences breeding success (Howard et al., 2024).

The scope of this thesis initially included a sensitivity analysis to identify the combination of parameters to best explain the migratory behaviour of *F. hypoleuca*. However, the thesis author was awarded funding to complete a PhD research project on using the SAMM to explain and predict *F. hypoleuca* migratory behaviour. Given additional funding, the scope of this thesis focused on methods to collect data on a migrants' arrival date and breeding success to help parameterise the terminal reward function within the model.

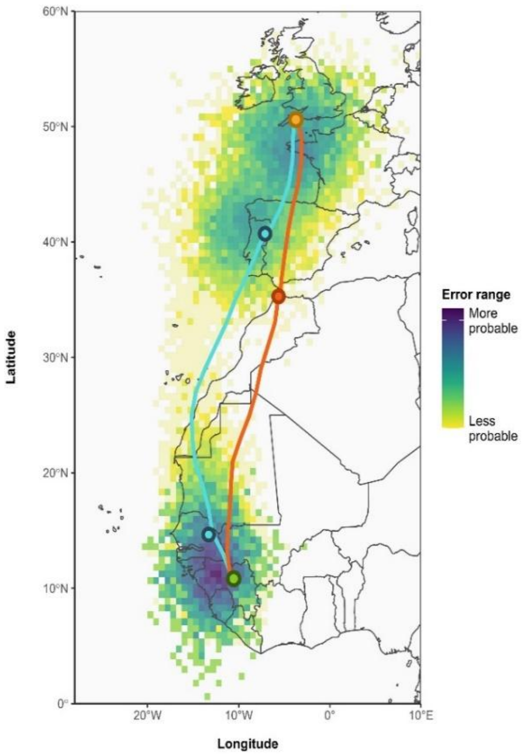

Figure 1.2, Geolocator data of the average migratory route of a British breeding population of F. hypoleuca during migration to and from breeding (orange) and non-breeding (green) grounds in autumn (blue) and spring (orange). With predominant stopover sites along the migration route. The figure was taken from Bell et al., 2022.

1.2 Challenges in Linking Migration to Reproductive Success

Researchers have used several approaches to relate time of arrival at the breeding ground with breeding success. These methods typically rely on tracking devices or colour ringing to record individual arrival in combination with nest box monitoring to relate arrival with breeding success (Bejarano & Jahn, 2018; Bell et al., 2024). In some cases, citizen science data have been used to estimate arrival dates at the population level (Koh & Opitz, 2025). However, these approaches present several limitations in producing a long-term and large-scale dataset on relating arrival dates to breeding success, including the costs of tracking technologies, which restrict the number of individuals that can be monitored; labour-intensive fieldwork requirements; and the need to either capture and mark individuals upon arrival or tag them during the previous breeding season - relying on their survival and site fidelity into the following year (Adamík et al., 2023; Bejarano & Jahn, 2018). This limitation was highlighted in a comprehensive analysis of *F. hypoleuca* arrival and first egg laying dates in the UK, which combined citizen science data for arrival dates with nest recordings schemes for egg-laying, wherein too few nests were followed to provide reliable metrics of breeding success (Nicolau et al., 2021).

Given these constraints, a key challenge in parameterising the terminal reward is collecting sufficient individual-level data on breeding phenology such as arrival dates and breeding success across populations in a manner that is both cost-effective and time-efficient.

1.3 Bioacoustics as a Tool for Scalable Monitoring

Faced with the time and cost constraints of traditional data collection methods, many ecological studies are increasingly adopting remote-sensing technologies to monitor species presence, abundance, and behaviour (Cavender-Bares et al., 2022; Reddy, 2021; Ross et al., 2023; Teixeira et al., 2024). These technologies often utilise advances in camera trapping and bioacoustics to enable non-invasive, scalable monitoring of individuals and populations (Delisle et al., 2021; Ross et al., 2023; Teixeira et al., 2024). A primary driver behind the growing use of bioacoustics is the advancement of audio processing software, such as BirdNET and Kaleidoscope Pro, that enable the analysis of large audio datasets by detecting audible behaviour, such as vocalisations (Gibb et al., 2019; Kahl et al., 2021; Pérez-Granados, 2023).

Acoustic monitoring is now readily deployed to monitor biodiversity (Alcocer et al., 2022), or target species distribution (Kramer et al., 2024), population density (Navine et al., 2024), and phenology (Van Doren et al., 2023). They are deployed across multiple habitats such as marine (Schall et al., 2024), freshwater (Desjonqueres et al., 2024), rainforests (Rappaport et al., 2022), grasslands (Shamon et al., 2021), urban areas (Levik et al., 2025), and mountain ranges

(Wood et al., 2023). They are also widely applicable to a range of taxa including fish (Souza et al., 2023), invertebrates (Raboin & Elias, 2019), mammals (Schall et al., 2024) and birds (van Merriënboer et al., 2024).

Within the study of bird migration, bioacoustics have been employed predominantly to detect vocalisations along flyways providing insights into migration phenology and intensity (Van Doren et al., 2023). With the expansion of bioacoustics, its application could be extended to capture finer-resolution data on a migrant's breeding phenology and breeding success. Frameworks developed to show how animal behaviour can be indicative of reproductive status (Berger-Tal et al., 2011), are now being applied to audible behaviours (Teixeira et al., 2019). So far, studies adopting this approach have focused on population-level breeding phenology in ground-nesting passerines (Schackwitz et al., 2020), and breeding success in burrow-nesting seabirds (Bennett et al., 2025) or at the nest-level in cavity-nesting cockatoos detecting occupancy and the success or failure of breeding attempts (Teixeira et al., 2022). In each case, estimates are based on a change in the rate of breeding-specific vocalisations over the breeding season.

For cavity-nesting passerines, such as *F. hypoleuca*, breeding stage phenology (e.g. onset of incubation, hatching, fledging) and breeding success (e.g. brood size) are typically estimated through traditional nest box monitoring (Walker et al., 2018). This approach is labour-intensive and time-consuming, leading to efforts to reduce the fieldwork through in-box camera technologies, which can track breeding stage onset and adult visitation rates (Hereward et al., 2021; Prinz et al., 2016; Zárybnická et al., 2016). However, these methods face limitations for large-scale application in migratory species, including hardware demands, disturbance from repeated access, and low flexibility to relocate devices to an active nest when target species are absent from the initial location (Hereward et al., 2021; Prinz et al., 2016).

Given the growing use of audio recorders for nest monitoring, their potential to detect audible behaviours at the nest box should be tested on cavity-nesting migrant passerines, such as *F. hypoleuca*. This could be achieved by positioning a recorder at a focal point, such as underneath a nest box, to detect audible behaviours attributable to the breeding pair. Changes in the frequency of these behaviours could provide information on breeding phenology (e.g. arrival at the nesting site or the onset of a breeding stage), breeding behaviour (e.g. nest box visitations), and reproductive outcomes (e.g. brood size). These estimates should then be compared with data from traditional nest box monitoring. If detections at the nest box provide reliable estimates of breeding phenology and success, acoustic monitoring may represent a promising non-invasive, passive, and continuous method for observing nest box activity. When paired with advanced processing tools, it could offer a cost-effective and scalable method for detecting the onset of breeding phenology stages such as arrival dates and measures of breeding success across individuals and populations (Gibb et al., 2019). Nonetheless, challenges remain, including the need for initial validation via complimentary monitoring, interference from environmental noise, and technical constraints such as battery life, data storage and processing requirements (Aide et al., 2013; Priyadarshani et al., 2018; Teixeira et al., 2019).

1.4 Objectives

This thesis tested the feasibility of detecting audible behaviours at the nest box, and whether changes in the frequency of detected behaviours can provide estimates for the onset of breeding stages (arrival at the nest box, onset of incubation, hatching, and fledging). Estimates compared with data collected from traditional nest box monitoring. If successful, this approach could provide a time- and cost-effective tool to collect individual-level breeding phenology and breeding success data to support the parameterisation of individual-based migration models.

Specific Objectives

1. To assess the feasibility of using audio recorders to detect breeding behaviours such as vocalisations (song and calls), and nest box visitations (entry/exit events)

2. To determine whether changes in the frequency of audible behaviours can be used to estimate breeding stage phenology, including arrival dates, onset of incubation, hatching, and fledging

3. To evaluate the relationship between detected behaviours and breeding outcomes, such as how parental visitation rates relate to chick growth and brood size.

4. To evaluate whether the breeding phenology and success data collected from nest box recordings can inform individual-based migration models.

1.5 References

Adamík, P., Bureš, S., Hahn, S., Oatley, G., & Briedis, M. (2023). Timing of migration and African non-breeding grounds of geolocator-tracked European Pied Flycatchers: a multi-population assessment. *Journal of Ornithology*, *164*(4), 875–886. https://doi.org/10.1007/s10336-023-02081-9

Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., & Alvarez, R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ*, *2013*(1). https://doi.org/10.7717/peerj.103

Alcocer, I., Lima, H., Sugai, L. S. M., & Llusia, D. (2022). Acoustic indices as proxies for biodiversity: a meta-analysis. *Biological Reviews*, *97*(6), 2209–2236. https://doi.org/10.1111/BRV.12890

Arlt, D., Olsson, P., Fox, J. W., Low, M., & Part, T. (2015). Prolonged stopover duration characterises migration strategy and constraints of a long-distance migrant songbird. *Animal Migration*, *2*(1), 47–62. https://doi.org/10.1515/ami-2015-0002

Barbet-Massin, M., Thuiller, W., & Jiguet, F. (2012). The fate of European breeding birds under climate, land-use and dispersal scenarios. *Global Change Biology*, *18*(3), 881–890. https://doi.org/10.1111/J.1365-2486.2011.02552.X

Barbet-Massin, M., Walther, B. A., Thuiller, W., Rahbek, C., & Jiguet, F. (2009). Potential impacts of climate change on the winter distribution of Afro-Palaearctic migrant passerines. *Biology Letters*, *5*(2), 248–251. https://doi.org/10.1098/rsbl.2008.0715

Bejarano, V., & Jahn, A. E. (2018). Relationship between arrival timing and breeding success of intra-tropical migratory Fork-tailed Flycatchers (Tyrannus savana). *Journal of Field Ornithology*, *89*(2), 109–116. https://doi.org/10.1111/jofo.12251

Bell, F., Bearhop, S., Briedis, M., El Harouchi, M., Bell, S. C., Castello, J., & Burgess, M. (2022). Geolocators reveal variation and sex-specific differences in the migratory strategies of a long-distance migrant. *Ibis*, *164*(2), 451–467. https://doi.org/10.1111/ibi.13017

Bell, F., Ouwehand, J., Both, C., Briedis, M., Lisovski, S., Wang, X., Bearhop, S., & Burgess, M. (2024). Individuals departing non-breeding areas early achieve earlier breeding and higher breeding success. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-53575-2

Bennett, S., Williamson, L., Hernández-González, M., Denton, T., Laber, R., Deakin, Z., Bolton, M., Manilow, E., & Wilson, L. J. (2025). Bioacoustics as a Measure of Population Size and Breeding Success of European Storm Petrels Hydrobates pelagicus. *Ecology and Evolution*, *15*(8). https://doi.org/10.1002/ECE3.71893

Berger-Tal, O., Polak, T., Oron, A., Lubin, Y., Kotler, B. P., & Saltz, D. (2011). Integrating animal behavior and conservation biology: A conceptual framework. *Behavioral Ecology*, *22*(2), 236–239. https://doi.org/10.1093/beheco/arq224

Berthold, P., Fiedler, W., Schlenker, R., & Querner, U. (1998). 25-Year Study of the Population Development of Central European Songbirds: A General Decline, Most Evident in Long-Distance Migrants. In *Correspondence to: P. Berthold Naturwissenschaften* (Vol. 85). Springer-Verlag.

Both, C., Bouwhuis, S., Lessells, C. M., & Visser, M. E. (2006). Climate change and population declines in a long-distance migratory bird. *Nature*, *441*(1), 81–83. https://doi.org/10.1038/NATURE04539;KWRD=SCIENCE

Buchan, C., Franco, A. M. A., Catry, I., Gamero, A., Klvaňová, A., & Gilroy, J. J. (2022). Spatially explicit risk mapping reveals direct anthropogenic impacts on migratory birds. *Global Ecology and Biogeography*, *31*(9), 1707–1725. https://doi.org/10.1111/geb.13551

Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., Fatoyinbo, L., Hurtt, G. C., Schimel, D., Townsend, P. A., Ustin, S. L., Wang, Z., & Wilson, A. M. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology and Evolution*, *6*(5), 506–519. https://doi.org/10.1038/s41559-022-01702-5

Clark, C. W., & Mangel, M. (2000). Dynamic State Variable Models in Ecology: Methods and Applications. *Dynamic State Variable Models in Ecology*. https://doi.org/10.1093/OSO/9780195122664.001.0001

Cowley, E., & Siriwardena, G. M. (2005). Long-term variation in survival rates of Sand Martins Riparia riparia: Dependence on breeding and wintering ground weather, age and sex, and their population consequences. *Bird Study*, *52*(3), 237–251. https://doi.org/10.1080/00063650509461397

Delisle, Z. J., Flaherty, E. A., Nobbe, M. R., Wzientek, C. M., & Swihart, R. K. (2021). Next-Generation Camera Trapping: Systematic Review of Historic Trends Suggests Keys to Expanded Research Applications in Ecology and

Conservation. In *Frontiers in Ecology and Evolution* (Vol. 9). Frontiers Media S.A. https://doi.org/10.3389/fevo.2021.617996

Desjonqueres, C., Linke, S., Greenhalgh, J., Rybak, F., & Sueur, J. (2024). The potential of acoustic monitoring of aquatic insects for freshwater assessment. *Philosophical Transactions of the Royal Society B*, *379*(1904). https://doi.org/10.1098/RSTB.2023.0109

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. In *Methods in Ecology and Evolution* (Vol. 10, Issue 2, pp. 169–185). British Ecological Society. https://doi.org/10.1111/2041-210X.13101

Goodenough, A. E. (2014). Effects of habitat on breeding success in a declining migrant songbird: The case of pied flycatcher ficedula hypoleuca. *Acta Ornithologica*, *49*(2), 157–173. https://doi.org/10.3161/173484714X687046

Goodenough, A. E., Elliot, S. L., & Hart, A. G. (2009). The challenges of conservation for declining migrants: Are reserve-based initiatives during the breeding season appropriate for the pied flycatcher ficedula hypoleuca? *Ibis*, *151*(3), 429–439. https://doi.org/10.1111/j.1474-919X.2009.00917.x

Hahn, S., Bauer, S., & Liechti, F. (2009). The natural link between Europe and Africa - 2.1 billion birds on migration. *Oikos*, *118*(4), 624–626. https://doi.org/10.1111/J.1600-0706.2008.17309.X;PAGEGROUP:STRING:PUBLICATION

Helm, B., Van Doren, B. M., Hoffmann, D., & Hoffmann, U. (2019). Evolutionary Response to Climate Change in Migratory Pied Flycatchers. *Current Biology*, *29*(21), 3714-3719.e4. https://doi.org/10.1016/j.cub.2019.08.072

Hereward, H. F. R., Facey, R. J., Sargent, A. J., Roda, S., Couldwell, M. L., Renshaw, E. L., Shaw, K. H., Devlin, J. J., Long, S. E., Porter, B. J., Henderson, J. M., Emmett, C. L., Astbury, L., Maggs, L., Rands, S. A., & Thomas, R. J. (2021). Raspberry Pi nest cameras: An affordable tool for remote behavioral and conservation monitoring of bird nests. *Ecology and Evolution*, *11*(21), 14585–14597. https://doi.org/10.1002/ece3.8127

Houston, A. I., & McNamara, J. M. (1999). *Models adaptive behaviour approach based state*. Cambridge University Press.

Howard, C., Mason, T. H. E., Baillie, S. R., Border, J., Hewson, C. M., Houston, A. I., Pearce-Higgins, J. W., Bauer, S., Willis, S. G., & Stephens, P. A. (2024).

Explaining and predicting animal migration under global change. *Diversity and Distributions*, *30*(2). https://doi.org/10.1111/ddi.13797

Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., Butchart, S. H. M., & Willis, S. G. (2020). Disentangling the relative roles of climate and land cover change in driving the long-term population trends of European migratory birds. *Diversity and Distributions*, *26*(11), 1442–1455. https://doi.org/10.1111/ddi.13144

Howard, C., Stephens, P. A., Tobias, J. A., Sheard, C., Butchart, S. H. M., & Willis, S. G. (2018). Flight range, fuel load and the impact of climate change on the journeys of migrant birds. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873). https://doi.org/10.1098/rspb.2017.2329

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

Koh, J., & Opitz, T. (2025). Extreme-value modelling of migratory bird arrival dates: insights from citizen science data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *188*(3), 674–699. https://doi.org/https://doi.org/10.1093/jrsssa/qnae108

Kramer, H. A., Kelly, K. G., Whitmore, S. A., Berigan, W. J., Reid, D. S., Wood, C. M., Klinck, H., Kahl, S., Manley, P. N., Sawyer, S. C., & Peery, M. Z. (2024). Using bioacoustics to enhance the efficiency of spotted owl surveys and facilitate forest restoration. *The Journal of Wildlife Management*, *88*(2), e22533. https://doi.org/10.1002/JWMG.22533

Kubelka, V., Sandercock, B. K., Székely, T., & Freckleton, R. P. (2022). Animal migration to northern latitudes: environmental changes and increasing threats. *Trends in Ecology & Evolution*, *37*(1), 30–41. https://doi.org/10.1016/J.TREE.2021.08.010

Levik, A., Dobromyslov, I., Goretskaia, M., Matasov, V., & Filyushkina, A. (2025). Ecoacoustic methods application for urban biodiversity monitoring. *Urban Ecosystems*, *28*(2), 1–17. https://doi.org/10.1007/S11252-024-01662-7/FIGURES/8

Nater, C. R., Burgess, M. D., Coffey, P., Harris, B., Lander, F., Price, D., Reed, M., & Robinson, R. A. (2023). Spatial consistency in drivers of population dynamics of a declining migratory bird. *Journal of Animal Ecology*, *92*(1), 97–111. https://doi.org/10.1111/1365-2656.13834

Navine, A. K., Camp, R. J., Weldy, M. J., Denton, T., & Hart, P. J. (2024). Counting the chorus: A bioacoustic indicator of population density. *Ecological Indicators*, *169*, 112930. https://doi.org/10.1016/J.ECOLIND.2024.112930

Newton, I. (2006). Can conditions experienced during migration limit the population levels of birds? *Journal of Ornithology*, *147*(2), 146–166. https://doi.org/10.1007/S10336-006-0058-4/TABLES/5

Nicolau, P. G., Burgess, M. D., Marques, T. A., Baillie, S. R., Moran, N. J., Leech, D. I., & Johnston, A. (2021). Latitudinal variation in arrival and breeding phenology of the pied flycatcher Ficedula hypoleuca using large-scale citizen science data. *Journal of Avian Biology*, *52*(2). https://doi.org/10.1111/jav.02646

Ouwehand, J., Ahola, M. P., Ausems, A. N. M. A., Bridge, E. S., Burgess, M., Hahn, S., Hewson, C. M., Klaassen, R. H. G., Laaksonen, T., Lampe, H. M., Velmala, W., & Both, C. (2016). Light-level geolocators reveal migratory connectivity in European populations of pied flycatchers Ficedula hypoleuca. *Journal of Avian Biology*, *47*(1), 69–83. https://doi.org/10.1111/jav.00721

Ouwehand, J., Asso, A. A., Johnston, B., Bot, S., Bil, W., Groenewoud, F., & Both, C. (2023). Experimental Food Supplementation at African Wintering Sites Allows for Earlier and Faster Fuelling and Reveals Large Flexibility in Spring Migration Departure in Pied Flycatchers. *Ardea*, *111*(1), 343–370. https://doi.org/10.5253/arde.2022.a37

Ozsanlav-Harris, L., Burgess, M. D., Hilton, G. M., & Bearhop, S. (2023). Changes in the spatial patterns of avian migrations: Evidence, mechanisms and causes. *Diversity and Distributions*, *29*(12), 1527–1545. https://doi.org/10.1111/ddi.13785

Pérez-Granados, C. (2023). BirdNET: applications, performance, pitfalls and future opportunities. *Ibis*, *165*(3), 1068–1075. https://doi.org/10.1111/ibi.13193

Prinz, A. C. B., Taank, V. K., Voegeli, V., & Walters, E. L. (2016). A novel nest-monitoring camera system using a Raspberry Pi micro-computer. *Journal of Field Ornithology*, *87*(4), 427–435. https://doi.org/10.1111/jofo.12182

Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology*, *49*(5), jav-01447.

Raboin, M., & Elias, D. O. (2019). Anthropogenic noise and the bioacoustics of terrestrial invertebrates. *Journal of Experimental Biology*, *222*(12). https://doi.org/10.1242/JEB.178749

Rappaport, D. I., Swain, A., Fagan, W. F., Dubayah, R., & Morton, D. C. (2022). Animal soundscapes reveal key markers of Amazon forest degradation from fire and logging. *Proceedings of the National Academy of Sciences*, *119*(18), e2102878119. https://doi.org/10.1073/PNAS.2102878119

Reddy, C. S. (2021). Remote sensing of biodiversity: what to measure and monitor from space to species? In *Biodiversity and Conservation* (Vol. 30, Issue 10, pp. 2617–2631). Springer Science and Business Media B.V. https://doi.org/10.1007/s10531-021-02216-5

Ross, S. R. P. J., O'Connell, D. P., Deichmann, J. L., Desjonquères, C., Gasc, A., Phillips, J. N., Sethi, S. S., Wood, C. M., & Burivalova, Z. (2023). Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. In *Functional Ecology* (Vol. 37, Issue 4, pp. 959–975). British Ecological Society. https://doi.org/10.1111/1365-2435.14275

Samplonius, J. M., Bartošová, L., Burgess, M. D., Bushuev, A. V., Eeva, T., Ivankina, E. V., Kerimov, A. B., Krams, I., Laaksonen, T., Mägi, M., Mänd, R., Potti, J., Török, J., Trnka, M., Visser, M. E., Zang, H., & Both, C. (2018). Phenological sensitivity to climate change is higher in resident than in migrant bird populations among European cavity breeders. *Global Change Biology*, *24*(8), 3780–3790. https://doi.org/10.1111/gcb.14160

Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., Woodruff, J., Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., & Woodruff, J. (2020). *Journal of Fish and Wildlife Management | www.fwspubs.org*. *11*. https://doi.org/10.3996/102019-JFWM-083

Schall, E., Kaya, I. I., Debusschere, E., Devos, P., & Parcerisas, C. (2024). Deep learning in marine bioacoustics: a benchmark for baleen whale detection. *Remote Sensing in Ecology and Conservation*, *10*(5), 642–654. https://doi.org/10.1002/RSE2.392

Shamon, H., Paraskevopoulou, Z., Kitzes, J., Card, E., Deichmann, J. L., Boyce, A. J., & McShea, W. J. (2021). Using ecoacoustics metrices to track grassland bird richness across landscape gradients. *Ecological Indicators*, *120*, 106928. https://doi.org/10.1016/J.ECOLIND.2020.106928

Shepard, E. L. C., Wilson, R. P., Rees, W. G., Grundy, E., Lambertucci, S. A., & Vosper, S. B. (2013). Energy landscapes shape animal movement ecology. *American Naturalist*, *182*(3), 298–312. https://doi.org/10.1086/671257

Shutt, J. D., Bell, S. C., Bell, F., Castello, J., El Harouchi, M., & Burgess, M. D. (2022). Territory-level temperature influences breeding phenology and reproductive output in three forest passerine birds. *Oikos*, *2022*(8). https://doi.org/10.1111/oik.09171

Souza, P. M., Olsen, Z., & Brandl, S. J. (2023). Paired passive acoustic and gillnet sampling reveal the utility of bioacoustics for monitoring fish populations in a turbid estuary. *ICES Journal of Marine Science*, *80*(5), 1240–1255. https://doi.org/10.1093/ICESJMS/FSAD085

Teixeira, D., Linke, S., Hill, R., Maron, M., & van Rensburg, B. J. (2022). Fledge or fail: Nest monitoring of endangered black-cockatoos using bioacoustics and open-source call recognition. *Ecological Informatics*, *69*, 101656. https://doi.org/10.1016/J.ECOINF.2022.101656

Teixeira, D., Maron, M., & van Rensburg, B. J. (2019). Bioacoustic monitoring of animal vocal behavior for conservation. *Conservation Science and Practice*, *1*(8), e72. https://doi.org/10.1111/CSP2.72

Teixeira, D., Roe, P., van Rensburg, B. J., Linke, S., McDonald, P. G., Tucker, D., & Fuller, S. (2024). Effective ecological monitoring using passive acoustic sensors: Recommendations for conservation practitioners. In *Conservation Science and Practice* (Vol. 6, Issue 6). John Wiley and Sons Inc. https://doi.org/10.1111/csp2.13132

Tomotani, B. M., van der Jeugd, H., Gienapp, P., de la Hera, I., Pilzecker, J., Teichmann, C., & Visser, M. E. (2018). Climate change leads to differential shifts in the timing of annual cycle stages in a migratory bird. *Global Change Biology*, *24*(2), 823–835. https://doi.org/10.1111/gcb.14006

Van Doren, B. M., Lostanlen, V., Cramer, A., Salamon, J., Dokter, A., Kelling, S., Bello, J. P., & Farnsworth, A. (2023). Automated acoustic monitoring captures timing and intensity of bird migration. *Journal of Applied Ecology*, *60*(3), 433–444. https://doi.org/10.1111/1365-2664.14342

Van Merriënboer, B., Hamer, J., Dumoulin, V., Triantafillou, E., & Denton, T. (2024). Birds, bats and beyond: evaluating generalization in bioacoustics models. *Frontiers in Bird Science*, *3*. https://doi.org/10.3389/fbirs.2024.1369756

Vriend, S. J. G., Grøtan, V., Gamelon, M., Adriaensen, F., Ahola, M. P., Álvarez, E., Bailey, L. D., Barba, E., Bouvier, J. C., Burgess, M. D., Bushuev, A., Camacho, C., Canal, D., Charmantier, A., Cole, E. F., Cusimano, C., Doligez, B. F., Drobniak, S. M., Dubiec, A., Eens M., Eeva T., Erikstad K., Ferns, P., Goodenough, A., Hartley, I., Hinsley S., Ivankina, E., Juskaitis, R., Kempenaers, B., Kerimov, A., Kalas, J., Lavigne, C., Leivits, A., Mainwaring, M., Martinez-Padilla, J., Matthysen, E., Van Oers, K., Orell, M., Pinxten, R., Reiertsen, T., Rytkonen, S., Senar, J., Sheldon, B., Sorace, A., Torok, J., Vatka, E., Visser, M., Sæther, B. (2023). Temperature synchronizes temporal variation in laying dates across European hole-nesting passerines. *Ecology*, *104*(2). https://doi.org/10.1002/ecy.3908

Walker, R. H., Robinson, R. A., Barimore, C. J., Blackburn, J. R., Barber, L. J., Clark, J. A., Clewley, G. D., Evans, H. E., Green, R. M., de Palacio, D. X., Grantham, M. J., Griffin, B. M., Schäfer, S., & Leech, D. I. (2018). Bird ringing and nest recording in Britain and Ireland in 2017. *Ringing and Migration*, *33*(2), 99–145.

Wang, X., Bell, F., Ouwehand, J., Bearhop, S., Burgess, M., & Both, C. (2022). *inherited trait variation during spring migration Tracking locally hatched pied flycatchers reveals masking of inherited trait variation during spring migration Abstract Climate change has been driving long-distance migratory birds to alter their schedule*.

Wittwer, T., O'Hara, R. B., Caplat, P., Hickler, T., & Smith, H. G. (2015). Long-term population dynamics of a migrant bird suggests interaction of climate change and competition with resident species. *Oikos*, *124*(9), 1151–1159. https://doi.org/10.1111/oik.01559

Wood, C. M., Champion, J., Brown, C., Brommelsiek, W., Laredo, I., Rogers, R., & Chaopricha, P. (2023). Challenges and opportunities for bioacoustics in the study of rare species in remote environments. *Conservation Science and Practice*, *5*(6), e12941. https://doi.org/10.1111/CSP2.12941

Zárybnická, M., Kubizňák, P., Šindelář, J., & Hlaváč, V. (2016). Smart nest box: A tool and methodology for monitoring of cavity-dwelling animals. *Methods in Ecology and Evolution*, *7*(4), 483–492. https://doi.org/10.1111/2041-210X.12509

Zimin, S. V., Zimin, A., Burns, D. J., Livne, R., Paz, R., Zvik, Y., Shochat, E., & Ovadia, O. (2023). Competition is a major limiting factor of refueling in migratory passerines during stopover. *Journal of Avian Biology*, *2023*(11–12). https://doi.org/10.1111/jav.03137

Zurell, D., Graham, C. H., Gallien, L., Thuiller, W., & Zimmermann, N. E. (2018). Long-distance migratory birds threatened by multiple independent risks from global change. In *Nature Climate Change* (Vol. 8, Issue 11, pp. 992–996). Nature Publishing Group. https://doi.org/10.1038/s41558-018-0312-9

2. A Proof-of-Concept for Acoustic Monitoring of Breeding Behaviour

Abstract

Traditional approaches to monitoring breeding behaviour are often labour-intensive, expensive, and intrusive. Bioacoustics has emerged as a promising alternative, but its capacity to detect breeding behaviours at nest boxes remains largely untested.

In this study, audio recorders were deployed outside nest boxes during breeding season, with detections validated against in-field observations. Custom classifiers were developed in Kaleidoscope Pro to detect vocalisations and visitation events, and their performance was assed using accuracy, precision, sensitivity, specificity and Matthews Correlation Coefficient. Detection frequencies were also compared with default BirdNET to evaluate the efficiency and reliability of using custom-classifiers over automated detection software.

Both vocalisations and visitation events were detectable. Classifiers performed best for song, while calls and visitation detections were less precise due to imbalanced datasets, which required greater manual validation. Comparisons with BirdNET suggested custom classifiers identified more events, though at the cost of additional development and processing time. Refining classifier performance and optimising recording schedules could improve the time efficiency of this approach.

Overall, this study provides a proof-of-concept that audio recorders represent a feasible and cost-effective method for monitoring nest box breeding behaviour, with potential for scaling across long-term monitoring networks.

2.1 Introduction

Monitoring breeding behaviour is central to understanding responses to environmental change and to broader population dynamics. Breeding timing, parental effort, and reproductive outcomes can vary across individuals in response to ecological conditions (Messmer et al., 2021; Porlier et al., 2012; Verhulst & Nilsson, 2007). Capturing this individual-level variation is essential for understanding how behavioural decisions relate to fitness and wider population trends, as well as informing conservation management (Porlier et al., 2012). However, acquiring such detailed data typically requires close and consistent monitoring of nesting individuals throughout the breeding cycle (Walker et al., 2018). Traditionally, this has involved active interventions such as in-person nest checks, colour-ringing (Walker et al., 2018; Wang & Beissinger, 2011), tracking technologies like radio-frequency identification (RFID) (Harrison & Kelly, 2022) and passive integrated transponder (PIT) tags (Nicolaus et al., 2008). These methods have produced high-resolution data on breeding phenology, nest attendance, provisioning rates, and reproductive output (Bejarano & Jahn, 2018; Harrison & Kelly, 2022; Koh & Opitz, 2025; Walker et al., 2018). However, they come with notable limitations: they are time-intensive, disruptive to individuals and nest sites (Nicolaus et al., 2008; Trolliet et al., 2014; Zárybnická et al., 2016), and often expensive to implement at scale (Iverson et al., 2023). These constraints highlight the need for less invasive, scalable monitoring tools capable of providing comparable ecological insight, either as complementary methods alongside traditional approaches or as independent alternatives where direct monitoring is impractical.

To overcome these constraints, researchers have adopted remote nest box monitoring tools, such as in-box camera systems. These allow for continuous behavioural observation with minimal disturbance, capturing events from nest construction through to fledging. However, commercially available nest box cameras are often limited in functionality. This has led to the development of custom-built cameras often using Raspberry Pi, which typically require box modifications or external housing for power and data storage. Such setups can cause disturbance to nests from repeated visits and offer limited flexibility if devices need to be relocated (Hereward et al., 2021; Prinz et al., 2016; Surmacki & Podkowa, 2022; Zárybnická et al., 2016). Even cost-conscious setups can exceed £200 per unit (Hereward et al., 2021), and access to dedicated funding for such hardware and associated processing software is limited.

Challenges in the scalability of custom-built nest box cameras highlight the need to explore alternative remote monitoring tools, such as external audio recorders, which may offer more affordable and less disruptive options for tracking breeding behaviour. Audio recording devices are now available in small, lightweight, and waterproof casings, enabling flexible attachment, access, and removal from the tree without direct disturbance to the individual or adaptation to the nest box.

Detecting nest-level behaviour using audio could include distinct audible cues such as adult vocalisations near the box, chick begging, and sounds associated with entry or exit events (Carlson et al., 2017). Advances in bioacoustics processing software, including tools such as Kaleidoscope Pro (capable of creating custom classifiers for target sounds) and BirdNET (representing an automated bird vocalisation detection when using default parameters), have enabled the easy processing of large acoustic datasets (Kahl et al., 2021; Wildlife Acoustics Inc, 2021). These tools have been shown to offer high precision (few false positives) in detecting target sounds, though challenges remain around sensitivity (missed detections), particularly in studies focused on behavioural changes over time wherein the frequency of detections is prioritised over presence-absence (Pérez-Granados, 2023).

Without visual confirmation, attributing acoustic events to specific individuals at the nest presents a challenge, requiring careful calibration and ground-truthing. While some sounds, such as entry and exit, can be confidently linked to the breeding pair through in-field validation, vocalisations may originate from nearby conspecifics. Therefore, developmental work is needed to assess whether audio recorders can reliably detect behaviours at nest boxes and whether processing requirements are feasible for broader application.

Objective

This chapter aimed to evaluate the feasibility of using audio recorders to detect key breeding behaviours at nest boxes by:

- **Identifying Behaviours of Interest:** detecting vocalisations (adult song and calls) and nest box activities (adult entry, exit, and chick begging).
- **Assessing Detection Range:** conducting a detection range assessment using degradation patterns in the signal-to-noise ratio to test if detections are specific to nest boxes.
- **Constructing Custom Classifiers**: constructing classifiers to detect vocalisation and visitation events.
- **Evaluating Classifier Performance**: Assessing the precision, sensitivity, and overall accuracy of each classifier using manually validated datasets.
- **Detection Method Assessment**: Comparing the performance of custom-built classifiers with established automated systems such as BirdNET to identify the most time-efficient and reliable method for behaviour detection.

2.2. Methods

2.2.1. Study Site and Species

The study was conducted within a woodland habitat in Snilesworth, North Yorkshire Moors, England (Lat-Lon: 54.3426, -1.2157), at an elevation of 150–170 m above sea level. The woodland is composed primarily of Oak (*Quercus robur*), Silver Birch (*Betula pendula*), and Alder (*Alnus glutinosa*), and is situated adjacent to the River Rye, providing a mix of riparian and deciduous forest habitats (Figure 2.1). Surrounding the woodland are areas of both arable and pastoral farmland. The woodland is located in a temperate climate, during the study period temperatures ranged from -1°C to 28°C with the mean maximum temperature 16°C and mean minimum temperature 7°C, and daily rainfall between 0 and 9 mm.

The study site is a known Pied Flycatcher (*Ficedula hypoleuca)* breeding site, with nest box monitoring having started in the late 1980s. The site contains approximately 125 nest boxes, which are used by a variety of UK woodland breeding bird species including Great Tit (*Parus major*), Blue Tit (*Cyanistes caeruleus*), Marsh Tit (*Poecile palustris*), Common Redstart (*Phoenicurus phoenicurus*), Eurasian Nuthatch (*Sitta europaea*), and Pied Flycatcher (*Ficedula hypoleuca*). The site typically supports approximately five to seven breeding pairs of *F. hypoleuca*; however, in the year of this study (2025), only three pairs were breeding on site and recorded. As a result, four *C. caeruleus* nests and two *P. Major* nests were also included in this study when assessing the ability to detect nest visitation at the nest box, *F. hypoleuca* remained the sole species studied for vocalisations.

Fieldwork was conducted from March 27th to June 12th 2025, spanning the period from the expected arrival of *F. hypoleuca* and encompassing the final fledging dates of all nestlings.

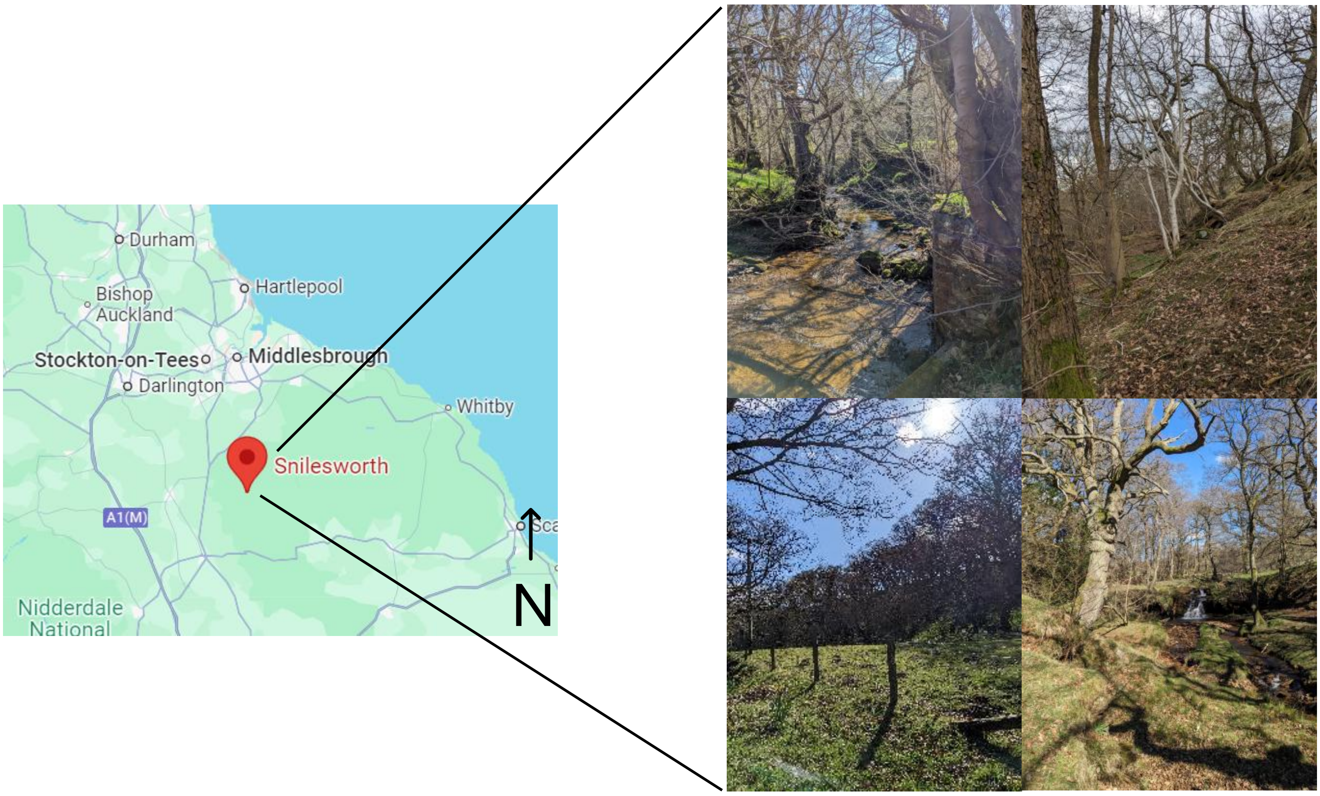

Figure 2.1 Study Site Location – Snilesworth, North Yorkshire Moors. The left panel shows the general location of the study site within the North York Moors in northeast England. The right panel presents a composite of site photographs, illustrating typical habitat features across the study area.

2.2.2. Recorder Distribution and Nest Box Occupancy

The distribution of recorders was informed by the nest boxes occupied by *F. hypoleuca* between 2018 and 2024 (Appendix A, Figure 2.A1). Across the 125 nest boxes at the study site, 22 nest boxes had been used, recorders were placed at 17 of these nest boxes with 3 other recorders placed to cover potential blind spots if any new boxes were to be occupied (Appendix A, Figure 2.A2).

Out of the 20 nest boxes with an audio recorder, six were initially occupied by either *P. major* or *C. caeruleus* (Figure 2.2.). No recorded nest boxes were initially occupied by the target species *F. hypoleuca* (Figure 2.2.). To detect *F. hypoleuca* vocalisations, audio data was taken from the nearest recorder, which in all three cases were between 20 to 40 meters of each other (Figure 2.3). Vocalisations were not detected for resident species (*P. major* or *C. caeruleus*) as the behaviour was focused on the sole migrant (*F. hypoleuca*) to estimate date of arrival at the breeding ground.

To enable the detection of visitations at the *F. hypoleuca* nest boxes, recorders were moved from an unoccupied recorded box during the incubation period (Figure 2.3). Once recorders had been moved to the *F. hypoleuca,* a total of nine nest boxes had audio recorders capable of detecting visitations at the nest box of *P. major, C. caeruleus,* and *F. hypoleuca* (Table 2.1). A full summary of the target behaviours for each recorder can be found in Table 2.1.

For the study site, a recorded nest box occupancy rate of 30% prior to moving the recorders is reasonable, as typical percentage range of nest boxes occupied at the site between 2018 and 2024 is 19% to 48%, with nest box occupancy in 2025 at 30% (Appendix A, Figure 2.A3). Since 2018, there have been seven species recorded at the study site, with *C. caeruleus* typically occupying 20 to 30 boxes, *P. major* showing moderate use at six to nine boxes, and *F. hypoleuca* occupying between two to seven nests per year (Appendix A, Figure 2.A4). Over the same period *F. hypoleuca* productivity was assessed from hatching success, with most nests producing six to eight chicks (the same range was recorded in 2025) but occasional failures and smaller broods recorded (Appendix A, Figure 2.A5). Across all nine recorded nest boxes, only one brood failed, all chicks in NB 22 died approx. 14 to 17 days post-hatching.

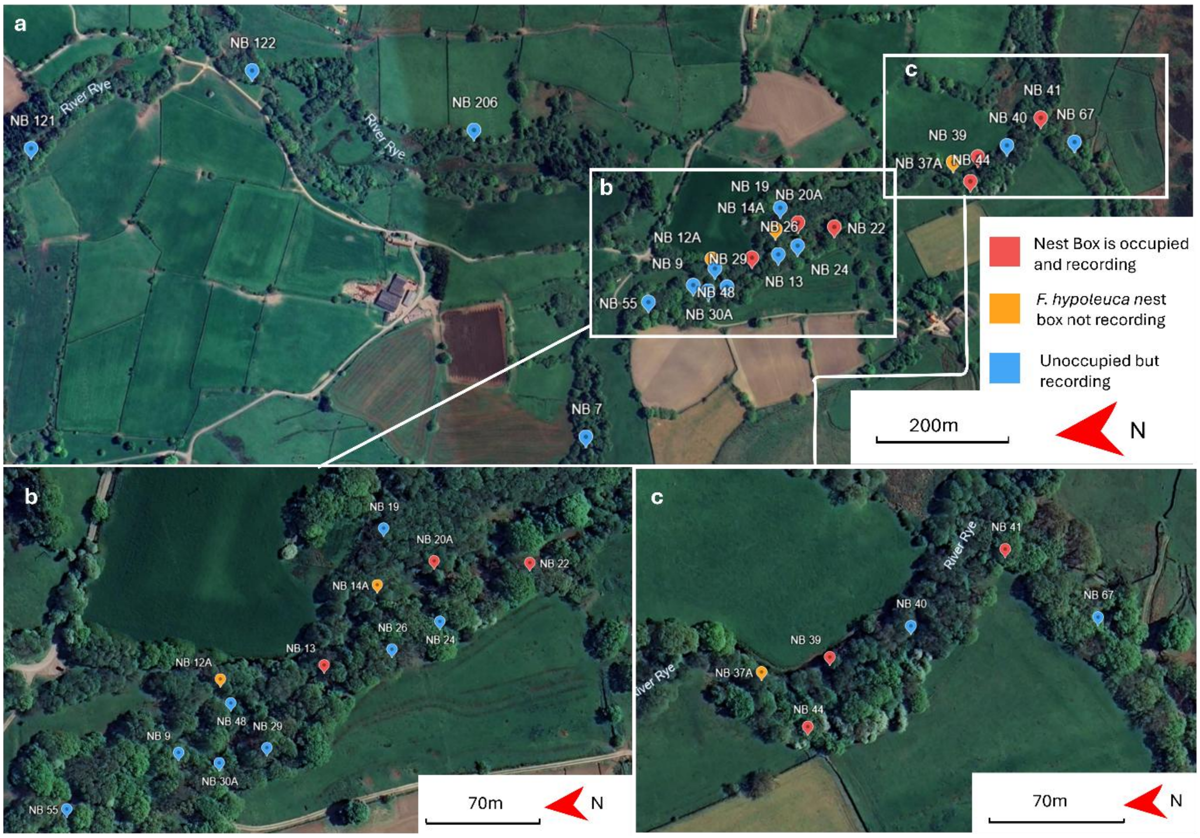

Figure 2.2, Map of study site showing the distribution of monitored nest boxes. (a) Overview of the study area with all nest boxes indicated. Insets (b) and (c) show enlarged views of two focal clusters of nest boxes. Colours denote nest box status: red = occupied and recorded, orange = Ficedula hypoleuca nest box not recorded, blue = unoccupied but recorded. Base map imagery from Google Earth 2025.

Figure 2.3, Map of recorder movement between nest boxes in relation to F. hypoleuca occupancy. Red markers indicate occupied F. hypoleuca nest boxes without a recorder, yellow markers indicate the nearest unoccupied recorded nest box, and orange markers indicate the nearest occupied recorded nest box. Arrows show the direction of recorder movement toward an occupied F. hypoleuca nest box. Prior to movement, vocalisations detected for F. hypoleuca were taken from audio data on the nearest recorder, i.e. NB48 for NB12A, NB19 for NB14A, & NB39 for NB37A. Base map imagery from Google Earth 2025.

Table 2.1, Summary of nest boxes, species, and behavioural targets monitored during the study. For F. hypoleuca, some nest boxes were recorded for both visitation and vocalisations, while others were recorded only for vocalisations prior to recorder movement to a F. hypoleuca nest box. Additional nest boxes occupied by P. major and C. caeruleus were monitored for visitation only.

Nest Box	Species	Target Behaviour
12A	*F. hypoleuca*	Both Visitation and Vocalisations
14A	*F. hypoleuca*	Both Visitation and Vocalisations
37A	*F. hypoleuca*	Both Visitation and Vocalisations
48	*F. hypoleuca*	Vocalisations (prior to recorder move to 12A)
19	*F. hypoleuca*	Vocalisations (prior to recorder move to 14A)
39	*F. hypoleuca*	Vocalisations (prior to recorder move to 37A)
13	*P. Major*	Visitation
39	*P. Major*	Visitation
20A	*C. caeruleus*	Visitation
22	*C. caeruleus*	Visitation
41	*C. caeruleus*	Visitation
44	*C. caeruleus*	Visitation

2.2.2. Recording Equipment and Configuration

Twenty Wildlife Acoustics Song Meter Micro 2 (SMM2) with built-in microphones were deployed at the study site. Recorders were positioned directly beneath nest boxes (Figure 2.4). Devices were securely mounted on trees, using locks in areas with easier public access (Figure 2.4a) and screws in less accessible locations (Figure 2.4b). On average, recorders were placed 35 cm below the nest box entrance hole and 277 cm above ground level. For each recorder, the microphone was positioned facing outward and not directed towards the nest box. Deployment occurred between 27 March and 4 April 2025. Each device recorded continuously after deployment, storing data in sequential 1-hour WAV files, at 24,000 Hz sample rate on a mono channel. Each recorder was set to 18 dB Gain, the device default, providing moderate sensitivity suitable for detecting vocalizations in a deciduous woodland environment.

Figure 2.4, Recording Equipment In-Situ. (A) Audio recorder secured to a tree using a locking strap. (B) Audio recorder mounted to a tree using screws. (C) Internal view of the recording device, showing component layout.

2.2.3. Audio Data Format and Management

Weekly visits were made to each recorder to replace the four nickel-metal hydride (NiMH) rechargeable batteries and to swap SD cards, which were either 64 GB or 128 GB in capacity (typical data usage per week of recording was 27GB) – Figure 2.4c. At each visit, the recorder's internal clock was logged to monitor for time drift and battery voltage was checked to ensure uninterrupted data capture. No instances of time drift were recorded throughout the monitoring period. During the first week of deployment, beginning 27 March, one recorder ceased recording due to battery depletion prior to the scheduled weekly check, resulting in a gap in data for that device. No other devices experienced recording failure throughout the monitoring period.

Files were stored in sequential 1-hour segments to facilitate data management and processing – producing approximately 168 WAV files per week. File metadata included recorder ID, date, time, and GPS coordinates.

2.2.4. Detection Range Assessment

To test whether recorded vocalisations were from individuals near the nest box, the detection range of audio recorders were assessed by measuring the signal-to-noise ratio (SNR) of test signals broadcast at known distances. SNR is the ratio between the amplitude of a target signal and the ambient noise, it is commonly used to assess sound degradation and detectability in field conditions (Araya-Salas et al., 2025; Araya-Salas & Smith-Vidaurre, 2017; Lampe et al., 2007). Three pure tones were generated using the WarbleR package in R (Araya-Salas & Smith-Vidaurre, 2017), based on the mean, minimum, and maximum frequencies of *F. hypoleuca* calls recorded at the site (rounded to 3800, 4500 and 6200 Hz). A 30 m transect was established extending from 5 m behind to 25 m in front of the recorder, with a clear direction of sight between audio and the SMM2. At 5m intervals, each tone was played once held at a height of 2.5m facing the recorder. The playback protocol was carried out on six randomly selected recorders, each receiving the full set of tones across 30m transect. Playback took place on 12 May 2025 between 1150 and 1440 under calm to light breeze conditions (0–7 mph, Beaufort scale) and at approximately 18 °C.

The SNR of each tone at every interval along the transect was calculated using the WarbleR package (Araya-Salas & Smith-Vidaurre, 2017) in R Studio (version 4.1.1.) (R Core Team 2021). Background noise amplitude was measured from the 0.2-second window immediately preceding each target signal. To ensure accurate SNR measurements, 8.7% of tones that overlapped with non-target bird vocalisations or environmental noise, were excluded.

To examine the change in SNR with distance and across the tested frequencies, a two-way ANOVA was conducted, with SNR as the response variable and distance and frequency as the explanatory variables. Post-hoc comparisons were performed using Tukey’s Honest Significant Difference (HSD) test.

2.2.5. Vocalisation and Visitation Clustering and Classifier Construction

To detect *F. hypoleuca* vocalisations at and visitations to the nest box, three tailored classifiers - one each for call, song, and adult visitations - were developed using Kaleidoscope Pro (Wildlife Acoustics Inc., 2021) . Classifier construction first requires a site-specific library of verified target signals. To build this library, a combination of methods was employed, including recording song and call in-situ, monitoring time of behaviour in-situ to later extract recording sample, and finally running a cluster analysis. The cluster analysis is a process that groups detections based on shared signal parameters such as frequency range, duration, and amplitude. This method facilitates the identification of acoustically similar sounds and the isolation of target signals. Once collated, the signal parameters are used to further improve the clustering of target sounds and exclude non-target sounds such as wind, rain, anthropogenic activity (e.g. vehicles), livestock, and non-target bird vocalisations. The classifiers are then constructed based on the site-specific library of target samples and the signal parameters. Once in-situ samples were identified, the Xeno-canto recordings were removed and not included in the call classifier development.

Site-specific Library of Vocalisations and Visitation Samples

The most effective method for developing a robust classifier was found to be the use of validated in-situ recordings of the target sound (Eberhardt, 2023). This approach necessitates the collection of in-situ recordings of *F. hypoleuca* calls, songs, and visitation. A diagram of the workflow to detect and collate a site-specific library for each target sound is provided in Figure 2.5.

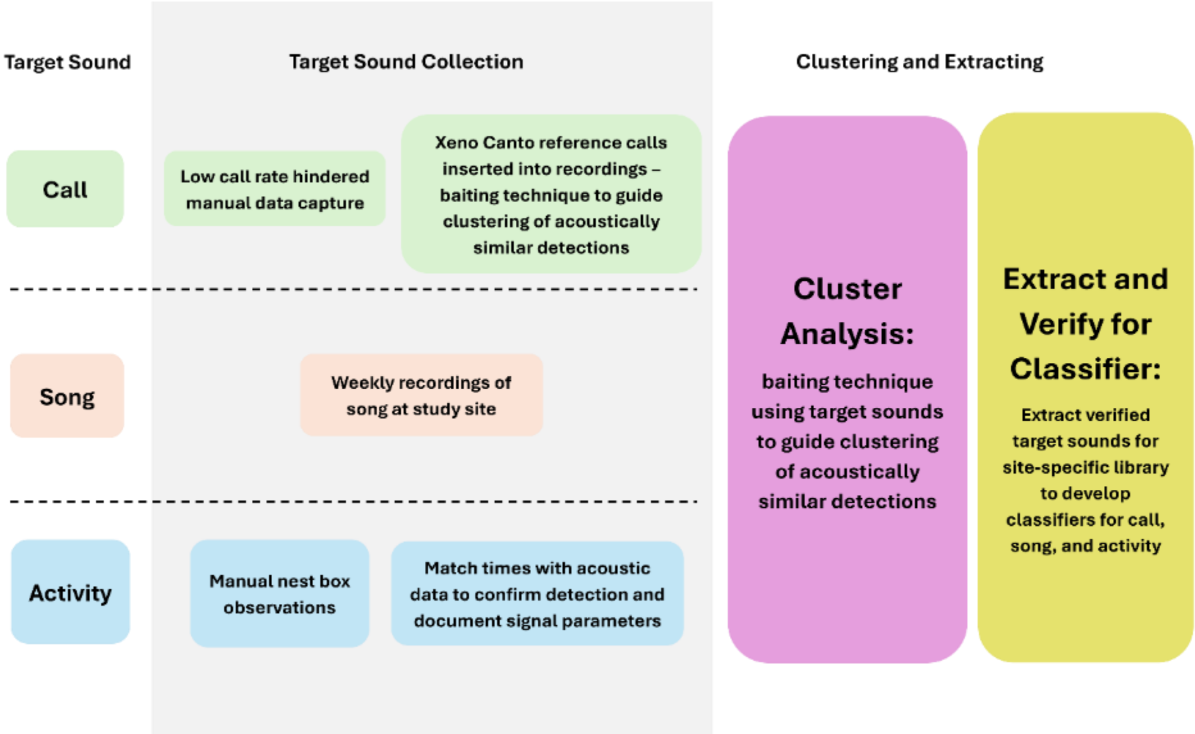

Figure 2.5, A schematic overview of the process used to identify and extract target vocalisations and visitation sounds to develop a site-specific library of target sound samples to be used in classifier construction.

Call Samples

Manual efforts to collect in-situ call recordings proved ineffective due to the infrequent production of calls during observation periods. To overcome this limitation, a baiting strategy was implemented to increase the number of sample recordings by inserting known *F. hypoleuca* calls sourced from UK populations via the Xeno-Canto database into the raw recording dataset. These reference calls guided the clustering process by attracting acoustically similar detections based on shared parameters such as frequency and duration. Clusters containing the reference calls were manually reviewed in Kaleidoscope's viewer, allowing for the identification of in-situ *F. hypoleuca* calls. This method resulted in 246 validated in-situ call samples to be used for the call classifier development.

Song Samples

For song, in-situ recordings collected during weekly visits to nest boxes provided sufficient material for cluster analysis. This process extracted a total of 460 target vocalisations from across the breeding period, to be used in the development of a dedicated song classifier.

Visitation Samples

In order to detect bird visitation, six nest boxes (two for each studied species; *F. hypoleuca, P. Major, and C. Caeruleus)* were monitored for approximately 1–2 hours between 0800 and 1200 within the first week post-hatching, when adult visitation is most frequent relative to nest building, laying, and incubation. During these sessions, the exact times of nest hole entry and exit events by adult birds were recorded manually. Each corresponding timepoint was then reviewed using Kaleidoscope Viewer to confirm that the visitation events were detected by the recorder. Once confirmed, the associated signal parameters were documented. The confirmed detections were seeded into new weekly raw data folders for the remainder of the nest being active (approx. 2 additional recording weeks) and initiate cluster analysis, enabling the extraction of 64 sample recordings. These samples facilitated the development of a classifier trained to detect nest visitation behaviour.

Target Signal Parameters

To assess the use of audio recorders for detecting *F. hypoleuca* behaviour at the nest box, the study focused on three target acoustic signals: song, call, and adult nest visitation. Figure 2.6 illustrates typical *F. hypoleuca* song, call, with adult nest visitation typical for all three study species.

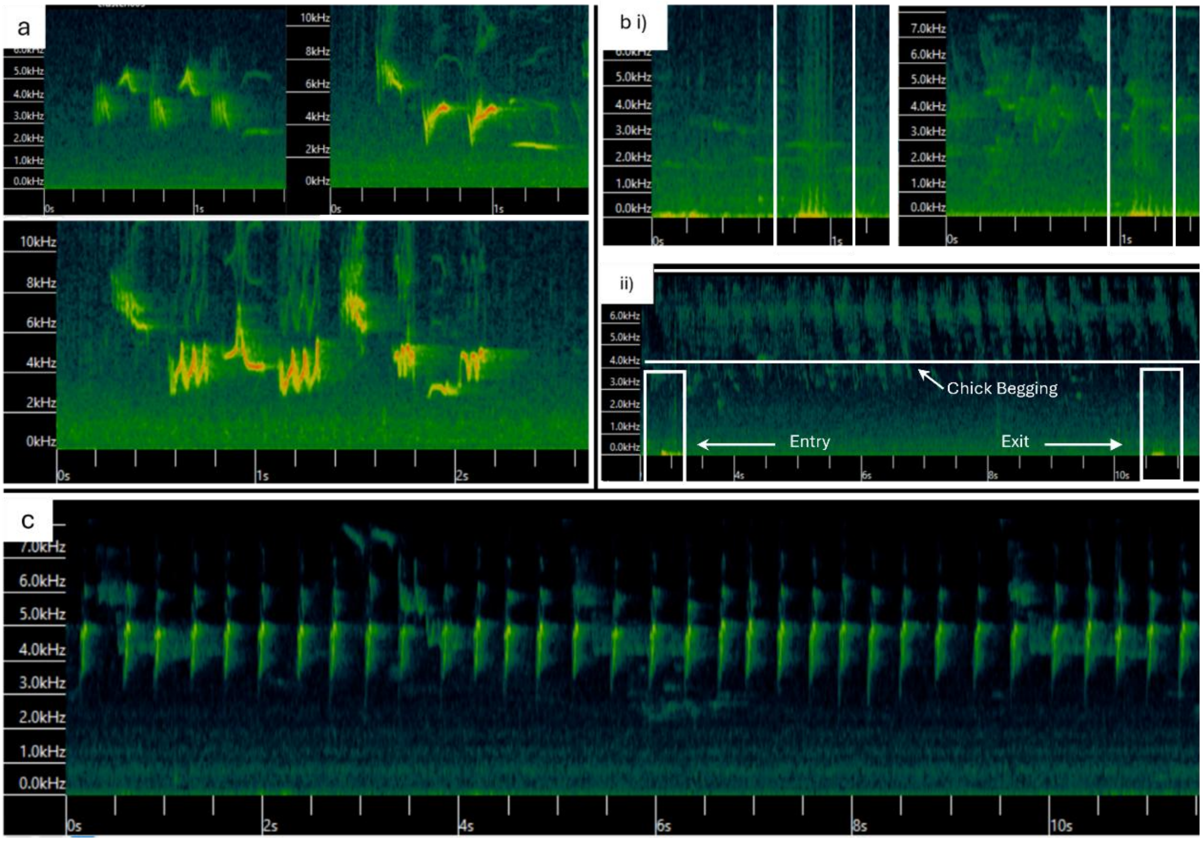

Figure 2.6, Spectrograms of Target Sounds. (a) Illustrates a subset of F. hypoleuca songs recorded at the study site, highlighting the variation and complexity in their structure. (b) Three examples of nest box activity detections, each with boundaries and/or annotations to highlight target behaviour. (bi) shows an adult entering or exiting the nest box, with wingbeats as the dominant acoustic feature. (bii) presents a combined detection of entry and exit activity, accompanied by chick begging calls immediately following adult arrival. (c) Examples of call vocalisations, these typically fall into two categories: alarm calls and simple contact calls. The example provided is characteristic of an alarm call. While calling rate may vary, these vocalisations are generally consistent in both duration and frequency range. Spectrograms display frequency in kilohertz (kHz) on the y-axis and time in seconds (s) on the x-axis. All images are screenshots captured using the Kaleidoscope Pro Viewer with the adjust contrast value -84.

Signal parameters associated with each target sound, such as frequency range, duration, inter-syllable gap, and amplitude, were initially assessed using in-situ recordings to define the acoustic boundaries of each category (Figure 2.7). These parameters were then used to constrain the search space during detection and clustering, increasing the likelihood of correctly identifying target sounds and reducing false positives from non-target vocalisations and environmental noise.

Table 2.2 summarises the average signal parameters derived for song, call, and visitation categories based on in-situ data collected at the study site. Figure 2.7 provides an example of a call signal; in this case, an alarm call, which typically produced a clearer recording due to proximity to the nest box. While some individual calls extended across a broader frequency range (c. 3200–6100 Hz, and in some cases with upper tails approaching 7000 Hz, as seen in Figure 2.7), the frequency range reported in Table 2.2 was restricted to the prominent signal to improve classifier performance by reducing noise and limiting overlap with non-target sounds.

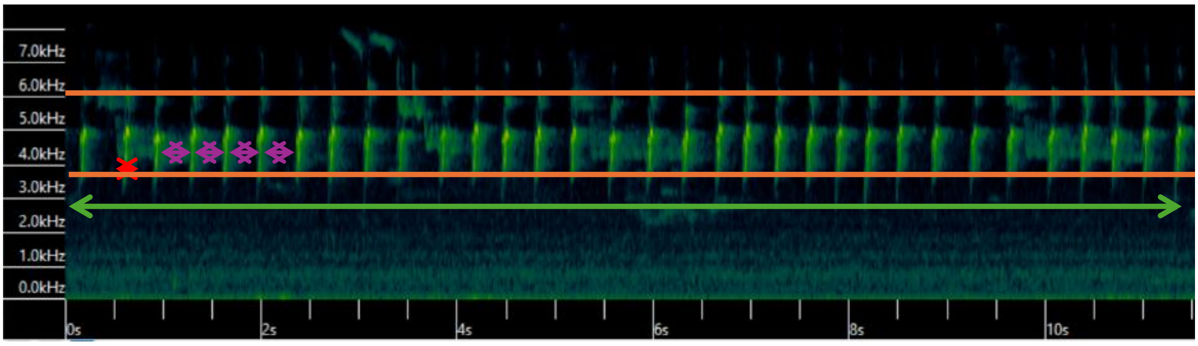

Figure 2.7, Diagram of measurements used to define signal parameters. Orange – frequency range, green – maximum detection length, red – minimum detection length, purple – maximum inter-syllable gap. Measurements were taken using Kaleidoscope Pro Viewer. Spectrograms display frequency in kilohertz (kHz) on the y-axis and time in seconds (s) on the x-axis. The signal shown in the spectrogram is an example of the F. hypoleuca alarm call.

Table 2.2 Signal parameters of target sounds. Summary of acoustic parameters used to define F. hypoleuca target vocalisations and visitations from all three study species. Frequency range is given in hertz (Hz), detection length in seconds (s), and maximum inter-syllable gap, also in seconds (s). Parameters are based on validated in-situ recordings.

Target Sound	**Frequency Range (Hz)**	**Length of Detection (s)**	**Maximum inter-syllable gap (s)**
Call	4000 - 6000	0.1 – 15	0.6
Song	2500 - 8000	0.1 - 15	0.35
Visitation	100 - 8000	0.1 - 15	0.6

Classifier Construction

Developing a classifier in Kaleidoscope Pro begins with a clustering analysis aimed at separating target sounds from non-target sounds. To quantify each detection's acoustic features, the software calculates Discrete Cosine Transform (DCT) coefficients from the sound spectrum. These coefficients provide measurable data used in the clustering process. Grouping of detections is performed using k-means clustering, guided by Hidden Markov Models (HMMs) derived from the DCT vectors (Pérez-Granados & Schuchmann, 2020). K-means clustering groups detections based on their distance from the cluster centre, which represents shared acoustic characteristics. The similarity threshold (default 0.5) controls how distinct clusters are from one another and therefore how much similarity is required for detections to be grouped together.

Once clusters are formed, the within-cluster threshold determines how far a detection can deviate from the cluster centre and still be classified as part of that cluster. This was set to the default value of 1, except for nest boxes 37A and 39. At these boxes, located near the river where acoustically similar low-frequency noise was elevated, the within-cluster threshold for the visitation classifier was lowered to 0.5. This required detections to more closely match the visitation cluster, thereby reducing the number of signals classified as visitation by excluding those with greater acoustic deviation. This adjustment was necessary to reduce the large number of detections (approximately 30,000), far exceeding the typical range (Table 2.3) likely as a result of background noises with similar signal parameters (Figure 2.10) classified as visitation producing false positives.

The clustering methodology enabled repeated extraction of target sound samples, which were reintroduced into the analysis as bait to refine the grouping process. This iterative approach improved clustering performance and supported the development of robust classifiers capable of consistently detecting the target sounds. For both song and call classifiers, four weeks of recordings (672 hours) were used for training, whereas for nest-box visitation classifiers, two weeks of recordings (one pre- and one post-hatching) from each of the three target species were used (1008 hours).

Once clusters were finalised, they are saved as a .KSC file, the classifier format used by Kaleidoscope Pro that stores the mathematical model of the clusters. Separate .KSC files were created for call, song, and visitation classifiers. Each classifier was then applied to all recordings from the study site to automate detection of the target sounds.

Data Processing

Each week of recorded data at a given nest box was processed independently using the three separate classifiers, enabling the detection and quantification of call, song, and visitation events across the breeding period. Using a classifier, a single analysis for 1 week's continuous recording (27GB) took approximately 10-15 minutes using a laptop with an AMD Ryzen 5 Pro 7535U processor (2.9 GHz) and 16 GB RAM, running a 64-bit operating system. To ensure false positives did not inflate the number of detections, each detection was reviewed manually. This task took approximately, 1 hour per 5,000 detections, with total weekly detections per recorder ranging from 279 to 15,818 – See Table 2.3.

Table 2.3 Summary of total target detections per week of recording for each classifier.

Classifier	Mean	Standard Deviation	Range
Call	1947	1079	419 - 5310
Song	5550	3006	279 - 11447
Visitation	3522	2524	328 - 15818

Classifier Performance

To test the performance of the developed classifiers, a subset of week-long recordings processed using each of the three classifiers were manually reviewed. For calls and song, one week-long recording from each of the three *F. hypoleuca* nest boxes was reviewed. For visitation, two week-long recordings were assessed for each of the three study species: *F. hypoleuca*, *C. caeruleus*, and *P. major*. These two weeks were selected to represent both pre-hatching and post-hatching periods, to account for changes in acoustic characteristics, such as the presence and intensity of chick begging calls.

Detections were assessed both audibly and visually using the Kaleidoscope Viewer by a single reviewer (AD). The same individual also recorded the in-situ samples, ran the clustering analysis, and constructed the classifiers, ensuring consistency across all stages of data collection and analysis. Each detection was evaluated both audibly and visually to determine whether it represented a target event (Table 2.4 – Actual Values). Classifier performance was then assessed by comparing these ground-truth labels with the classifier's predictions. Based on this comparison, detections were categorised into one of four outcomes: true positives (correctly identified target sounds), false positives (non-target sounds misclassified as target sounds), false negatives (missed target sounds), and true negatives (correctly identified non-target sounds).

Table 2.4 Confusion matrix framework illustrating possible outcomes for classifier predictions compared to actual values. Green cells indicate a correct classification, red cells indicate an incorrect classification.

		Actual Values	
		Positive	**Negative**
Predicted Values	**Positive**	True Positive	False Positive
	Negative	False Negative	True Negative

The manual review was used to derive standard performance metrics, including precision, sensitivity, specificity, accuracy and Mathews Correlation Coefficient (Table 2.5), for each of the three classifiers (Eberhardt, 2023; Enari et al., 2019; van Merriënboer et al., 2024). The Matthews Correlation Coefficient (MCC) was chosen as it provides a more balanced evaluation when target and non-target sounds are imbalanced (van Merriënboer et al., 2024), as is often the case in ecological datasets (Mellor et al., 2015).

Table 2.5 Summary of performance metrics for each classifier, including the formula used for calculation and the purpose of each metric in evaluating classifier performance. TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative, N = Total Negatives (FP + TN).

Metric	Calculation	Purpose
Precision	$\frac{TP}{TP + FP}$	Indicates how many identified targets are correct
Sensitivity (Recall)	$\frac{TP}{TP + FN}$	Measures how many actual target sounds were correctly identified
Specificity	$\frac{TN}{TN + FP}$	Measures how well non-target sounds were correctly ignored
Accuracy	$\frac{TP + TN}{N}$	Overall proportion of correct classifications (target and non-target)
Matthews Correlation Coefficient (MCC)	$\frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	Balanced metric reflecting overall performance, even with class imbalance. Scale (-1 to 1)

2.2.6. Imbalance Ratio

Class imbalance is the difference between the number of non-target and target sounds (Saito & Rehmsmeier, 2015). In ecology this imbalance is particularly common when the presence of relatively rare classes (such as target vocalisations and visitation) occurs far less frequently in comparison to other classes (non-target sounds) (Mellor et al., 2015). When such imbalances are present, they can distort performance metrics. For example, accuracy (the proportion of correct predictions) can appear deceptively high. In a highly imbalanced dataset, a classifier that simply labels every instance as "non-target" might achieve high accuracy by correctly identifying the majority class, despite failing entirely to detect rare but important target events. This inflation of true negatives (TN) and overall negatives (TN + FP) masks poor sensitivity and renders accuracy a misleading indicator of true performance. In such cases, focusing on metrics such as precision, sensitivity, and MCC are more reliable (van Merriënboer et al., 2024). To quantify this effect, Equation 1 was used to calculate the imbalance ratio for each of the three classifiers, based on the same data used to construct their respective confusion matrices. The imbalance ratio was then used to assess the reliability of performance metrics, helping to prevent over-reliance on potentially misleading measures such as accuracy in the presence of an imbalanced dataset.

Equation 1:

$$Imbalance\ Ratio = \frac{No. Non - Target\ Sounds}{No. Target\ Sounds}$$

2.2.7. Detection Method Comparison

To compare the performance of two automated acoustic detection tools, BirdNET and Kaleidoscope, this study assessed their ability to detect *F. hypoleuca* vocalisations across the recorded period. For Kaleidoscope, song and call detections using their respective classifiers were combined to cover all vocalisations. For BirdNET, each recording was processed using BirdNET

Analyser (Kahl et al., 2021). Kaleidoscope represents a site-specific and relatively time-intensive method for acoustic detection, whereas BirdNET offers a fully automated and generalised approach requiring minimal manual input. BirdNET processed the recordings at a rate of 25 seconds per 1 hour file, with all recordings from a single nest box recorder processed within approx. 10 hours. Compared to an approximate total processing time of 15 hours per nest box recorder when using Kaleidoscope Pro. Detections equal or greater than 0.75 confidence were assumed to be true positive detections (Pérez-Granados, 2023).

Detections were binned into 15-minute fixed time intervals to enable direct comparison between methods. The presence or absence of detections within each time bin was used to quantify overlap and disagreement. A categorical variable was created per time bin; “Both” – detected by both methods, "BirdNET only" or "Kaleidoscope only", and “No Detection”. For each nest box, Cohen’s Kappa was used to evaluate the level of agreement between the detection methods (Kahl et al., 2021), and McNemar’s test was used to evaluate if the disagreement between the two methods was unbalanced i.e. is one method more likely to make a detection when the other does not (Pembury Smith & Ruxton, 2020).

Each test was also performed using 5-, and, 10- time bins to ensure results were robust to changes in temporal resolution. Actograms (Showing daily patterns of visitation across the breeding period) were constructed to illustrate temporal detection patterns for each nest box. All analysis and visualisations were conducted in R Studio (version 4.1.1.) (R Core Team 2021).

2.3. Results

2.3.1 Detection Range Assessment

There was a significant effect of distance on signal-to-noise ratio (SNR) (ANOVA: F(7,105) = 29.81 $p<0.001$). However, there was no significant effect of frequency (3.8, 4.5, 6.2 KHz) on SNR (ANOVA: F(2,105) = 0.99 $p>0.1$).

Post-hoc Tukey's HSD tests showed that SNR was significantly higher at closer distances (1m and 5m) compared to farther distances (15–25 m) and background recordings. Notably, there were no significant differences in SNR between distances of 15 m, 20 m, 25 m, and background. The sound degradation profiles along the transect and across tested frequencies are illustrated in Figure 2.8. Full output tables from Tukey HSD test can be found in Appendix A Table 2.A1.

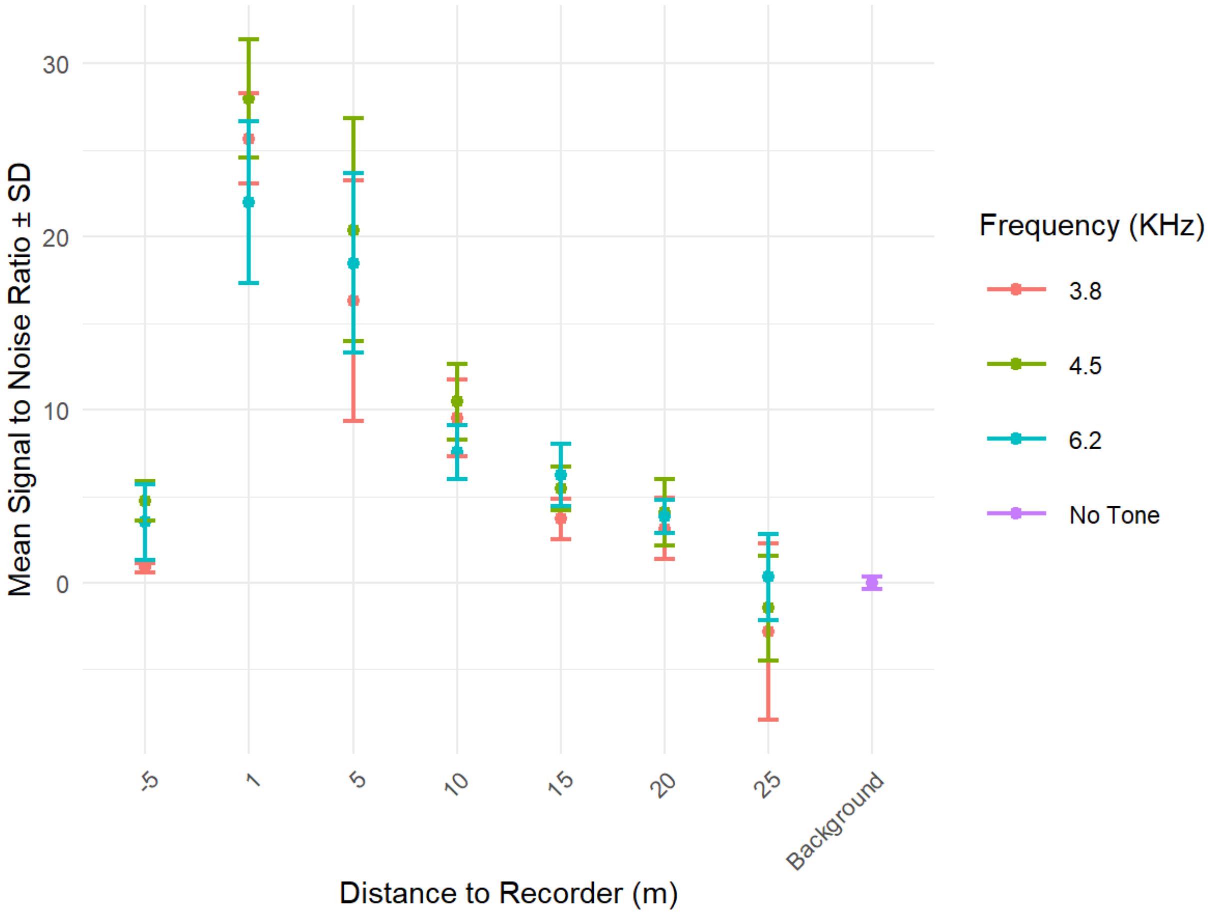

Figure 2.8 Mean Signal to Noise Ratio (SNR) degradation profile along the playback transect in meters (m) for three test frequencies 3.8, 4.5, 6.2 kilohertz (KHz). Mean Signal to Noise Ratio and standard deviation values (SD) are taken from the six recorders included in this test. Negative distances to the recorder (-5) are measurements taken from behind the tree with the recorder and nest box attached, positive values are taken along the transect in front of the recorder. Background is used as a control, wherein no tone was played, instead using a sample prior to playback was analysed.

2.3.2. Kaleidoscope Pro Classifier Performance Metrics

Kaliedoscope classifier performance varied across the three target detection tasks (call, song, visitation), as shown in Table 2.6. The *F. hypoleuca* call classifier exhibited high sensitivity (0.902 ± 0.094), indicating that it successfully identified most true call events. However, precision was low (0.199 ± 0.269), reflecting a high false positive rate (FPR = 0.045 ± 0.032), which suggests frequent misclassification of non-target sounds as calls. This is further reflected in its modest Matthews Correlation Coefficient (MCC = 0.315 ± 0.296), which indicates performance is being pulled down by the imbalance between sensitivity and precision.

The *F. hypoleuca* song classifier showed the most balanced performance, with similar values for sensitivity (0.640 ± 0.032) and precision (0.648 ± 0.239), and the highest MCC (0.568 ± 0.105), indicating good discriminative ability. The relatively low false positive rate (0.058 ± 0.010) and moderate accuracy (0.855 ± 0.036) suggest effective separation of target and non-target song events.

The visitation classifier used across all three study species (*F. hypoleuca, C. caeruleus,* & *P. major)* achieved the highest overall accuracy (0.965 ± 0.009) and specificity (0.966 ± 0.009), reflecting strong rejection of non-target detections. However, its precision remained low (0.136 ± 0.118) despite reasonable sensitivity (0.774 ± 0.154), likely due to misclassification of background low-frequency noise as nest visitation. The resulting MCC (0.286 ± 0.152) reflects, similar to the call classifier, a trade-off between sensitivity and precision driven by class imbalance. As shown in Table 2.6, the visitation classifier faced a markedly higher non-target to target ratio (480:1), which inflated true negatives and lowered MCC. By comparison, the song classifier faced a more balanced ratio.

Table 2.6, Summary of performance metric results for each of the three classifiers. The mean coefficient is provided with the standard deviation in brackets. Results are colour coded by classifier performance with red indicating values below 0.25, orange indicating values between 0.25 and 0.75, and green indicating values above 0.75.

	Precision	**Sensitivity**	**Specificity**	**Accuracy**	**MCC**
Call	0.199 (+/- 0.269)	0.902 (+/- 0.094)	0.955 (+/- 0.035)	0.954 (+/- 0.031)	0.315 (+/- 0.296)
Song	0.648 (+/- 0.239)	0.64 (+/- 0.032)	0.942 (+/- 0.009)	0.855 (+/- 0.036)	0.568 (+/- 0.105)
Visitation	0.136 (+/- 0.118)	0.774 (+/- 0.154)	0.966 (+/- 0.009)	0.965 (+/- 0.009)	0.286 (+/- 0.152)

2.3.3. Imbalance Ratio

As shown in Table 2.7, the imbalance ratio is relatively high for all three classifiers, and exceptionally high for the call and visitation classifiers - indicating that for every one target sound, there are approximately 4,553 non-target sounds in call, 480 in visitation, and 7 in song. Given these substantial imbalances, interpretation of the classifier performance results should prioritise metrics that are less sensitive to class distribution, such as precision, sensitivity, and the Matthews Correlation Coefficient (MCC).

Table 2.7, Summary of the imbalance ratio calculated for each of the three classifiers.

Target Sound	Imbalance Ratio
Call	4553
Song	7
Visitation	480

2.3.4. Detection Method Comparison

When detections between 15-minute time intervals were compared, agreement between Kaleidoscope and BirdNET detections was generally good across the three nest boxes (Table 2.8). Cohen's Kappa values ranged from 0.45 – 0.75, indicating that both methods are flagging detections in the same time intervals at a higher rate than would be expected randomly. Despite this overall agreement, McNemar's test was highly significant for all boxes ($p < 0.001$, Table 2.10), showing that disagreements were systematically unbalanced, with one method more likely than the other to record a detection when the other did not. For nest boxes 14A and 37A (Figure 2.2), Kaleidoscope detected substantially more vocalizations than BirdNET (Table 2.8). In contrast, in nest box 12A (Figure 2.2), BirdNET detected 465 time-bins that were missed by Kaleidoscope, compared to only 204 missed by BirdNET. Results from tests using 5- and 10-minute time showed no meaningful differences from the 15-minute bin results reported here and in Table 2.8, indicating that the detection patterns and agreement metrics are robust to changes in time bin resolution.

Visual comparisons of the detection patterns for each method are shown in Figure 2.9. Typically, agreement between methods occurs during the day, with Kaleidoscope outperforming BirdNET later in the day, while BirdNET is detecting vocalisations earlier in the day. For Nest Box 12A, the band of predominantly BirdNET detections between 16th May and 20th May does not align with any recorder interference from in-person monitoring, nor an absence of recording effort, repeated tests of this period using the Kaleidoscope classifier did not yield different results. The reason for this discrepancy remains unclear, though it may relate to classifier sensitivity to the wider soundscape or recording conditions during this period.

Table 2.8, Comparison of detection performance between Kaleidoscope Pro and BirdNET across entire breeding period for the three F. hypoleuca nest boxes. Outputs from Cohen's Kappa and McNemar analyses are shown, along with the number of overlapping time intervals with detections by both methods, number of intervals with detections for each method, and number of intervals with no detections by either method.

Nest Box	Cohen's Kappa (95% CI)	McNemar X^2 (p)	Detected by both methods	BirdNET only	Kaleidoscope only	No detections by either method
14A	0.64 (0.62, 0.66)	731.25 ($p < 0.001$)	1645	144	744	2523
37A	0.45 (0.43, 0.47)	1137.4 ($p < 0.001$)	975	55	1308	3351
12A	0.75 (0.73, 0.77)	101.05 ($p < 0.001$)	2519	459	211	2212

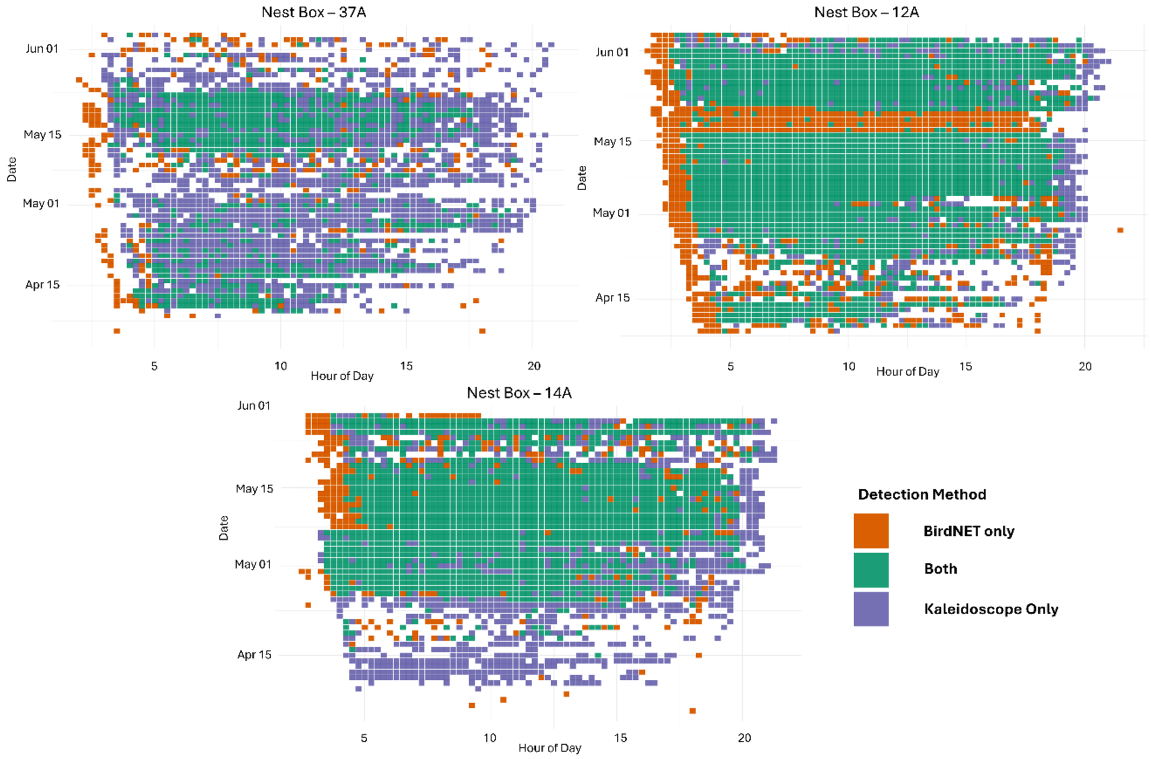

Figure 2.9, Detection Overlap Assessment. Actograms showing the overlap in detections made by Kaleidoscope Pro and BirdNET across the study period. Each row represents a day of recording, binned into 15-minute intervals, with hour of day plotted on the x-axis. Shaded bins indicate periods of detection by each method: red for BirdNET, purple for Kaleidoscope Pro, and green for overlapping detections. Overlap suggests agreement between methods, while non-overlapping detections reflect potential differences in sensitivity or classification. Each actogram illustrates one of three F. hypoleuca nest boxes, NB12A, NB14A, and NB37A.

2.4 Discussion

2.4.1. Ability to Detect Audible Behaviour at the Nest Box

The results demonstrate that a range of audible behaviours at the nest box, particularly vocalisations and visitations, can be reliably detected using passive audio recorders. This confirms the potential of acoustic monitoring as a non-invasive tool for capturing fine-scale behavioural data at the nest box, with detectable behaviours including adult vocalisations, nest box entry and exit events, and chick begging.

2.4.2. Classifier Performance and Vocalisation Detection

The classifiers developed in Kaleidoscope Pro using both validated recordings and cluster analysis outputs were clearly capable of detecting vocalisations at the nest box. However, the relatively low sensitivity of song detection (i.e. ability to capture all target sounds), highlights ongoing challenges in designing classifiers that capture the full diversity of target species' vocalisations, given natural variation in call types, amplitude, and background noise conditions (Marsland, 2011; Priyadarshani et al., 2018).

Alternatively, the use of fully automated software such as BirdNET may offer a faster and more generalised approach by drawing on a broader library of non-site-specific vocalisations to inform detection and acoustic data processing (Kahl et al., 2021).

In comparing the site-specific Kaleidoscope classifiers with BirdNET, the site-specific classifiers had improved detection rates for two of the three nest boxes (Table 2.7), particularly in acoustically complex environments. For instance, at nest box 37A, located near a river, Kaleidoscope-only detections far exceeded BirdNET only detections (Figure 2.9 & Table 2.7), likely due to its exposure to recordings including low-frequency interference during model training. These findings suggest that the choice of detection method should be aligned with study aims. Where the objective is to monitor behavioural changes over time, especially in variable soundscapes, site-specific classifiers such as those generated in Kaleidoscope are likely more reliable. However, the added processing time and manual calibration required may limit feasibility for large-scale deployments (Pérez-Granados, 2023).

2.4.3. Nest Visitation Detection and Acoustic Interference

Beyond adult vocalisations where the sole effort was to identify *F. hypoleuca* call and song, adult nest box visitations were also clearly detectable across the three species *F. hypoleuca, C. caeruleus,* and *P. major*. These were typically characterised by distinct wingbeat sounds, movement noise, or chick vocalisations triggered by adult arrival. Although directionality (entry vs exit) could not be distinguished from audio alone, this limitation is shared with other methods such as RFID (Harrison & Kelly, 2022).

Spectrograms showed that visitation events often began with low-frequency bursts, rising in frequency and tapering in amplitude – features that are shared with background noises such as wind or nearby water flow (Figure 2.10). This overlap is evident at two nest boxes during the study, where proximity to running water led to inflated false positive detections and increased manual validation processing time.

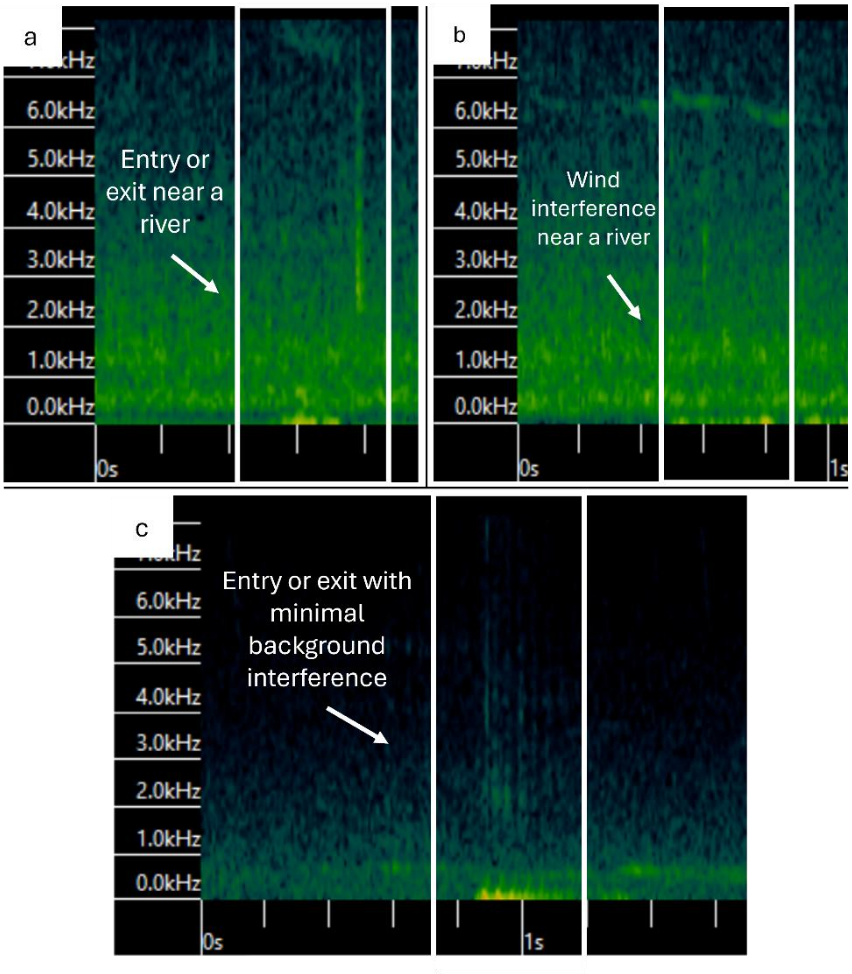

Figure 2.10, Comparison between nest visitation detections and interference from the environment. Panel a is an example of entry or exit from the nest box in the presence of low-frequency background noise from a nearby river, whereas panel b is an example of wind interference from the same recorder. When compared, their similarities in frequency and duration highlight the potential confusion during automated detection. For reference, panel c is a detection of a nest box entry/exit event recorded under minimal background interference, illustrating a clearer signal profile. Spectrograms display frequency in kilohertz (kHz) on the y-axis and time in seconds (s) on the x-axis. All images are screenshots captured using the Kaleidoscope Pro Viewer, with the adjust contrast value -84.

As the breeding season progressed, the acoustic characteristics of visitations changed with breeding stage. In particular, chick begging vocalisations became more prominent as nestlings developed, producing louder and more distinct sounds upon the arrival of adults, as illustrated in Figure 2.11. These higher-amplitude vocalisations may provide a clearer acoustic target against consistent low-frequency background noise, potentially improving detectability during later breeding stages. If so, this shift could serve as a distinct and detectable marker of breeding behaviour, offering studies of breeding phenology that are currently relying on citizen science data, a novel phenological marker (Massimino et al., 2021; Newson et al., 2016).

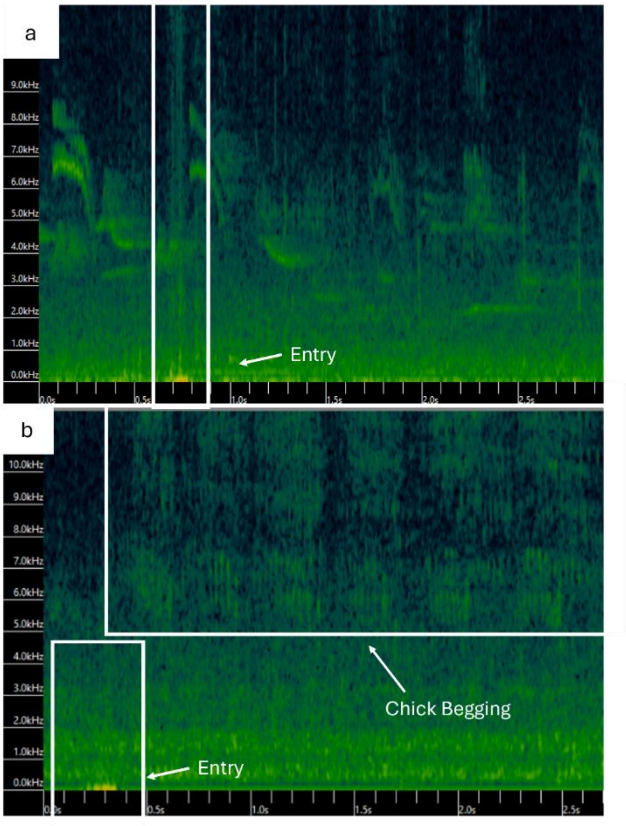

Figure 2.11, Comparison of nest box entry or exit events recorded adjacent to the river. Panel a shows an entry or exit detection prior to the onset of chick begging, while panel b shows a detection after the onset of chick begging. The panels illustrate the change in signal structure across the breeding timeline and how this may affect the detectability of adult visitation. Spectrograms display frequency in kilohertz (kHz) on the y-axis and time in seconds (s) on the x-axis. All images are screenshots captured using the Kaleidoscope Pro Viewer with the adjust contrast value -84.

2.4.4. Detection Range and Attribution of Vocalisation

One objective of this study was to evaluate whether nest box recorders could capture behaviours at the target nest. While the detection of visitation events confirmed the ability of tracking adult visitation at the box, attributing vocalisations to specific individuals proved more complex. Vocalisations, unlike visitations, are not spatially constrained and may originate from nearby conspecifics.

Sound degradation tests confirmed that detection range decreased with distance from the source, with a potential detection radius of approximately 25m for the *F. hypoleuca* vocalisation frequency range (Figure 2.8). This is notably shorter than the 80m range estimated by Winiarska et al., (2024) who used higher microphone placements (3-9m) and sampled across a broader seasonal window. In contrast, this study set the speakers at 2.5 m to play back the test tones at the average height of the nest box and recorder and conducted the trials in mid-May, conditions that likely reduced the effective range due to seasonal changes in sound attenuation from differences in vegetation structure at different heights (Gaudon et al., 2022).

Ultimately, the reliability of detecting individual vocalisations at the nest box depends on sound degradation patterns and the density of the target species near the nest site (Knight et al., 2024; Winiarska et al., 2024). Given the distribution of *F. hypoleuca* nests at the study site, with nearest-neighbour distances exceeding 80m, such overlap is unlikely to pose a significant issue in this context. This uncertainty may pose challenges at study sites with high nesting densities, where overlapping detection ranges complicate individual identification. In such contexts, studies prioritising individual-level behavioural observation may benefit from focusing on detecting visitations over vocalisations. Although, models are increasingly capable of navigating around these issues by differentiating between individuals based on variation in vocalisation structure (Knight et al., 2024; Stowell et al., 2019). For now, with calibration for site-specific acoustic conditions, nest box–mounted audio recorders offer a viable method for detecting behaviours at and in the vicinity of the nest.

2.4.5. Feasibility of Detecting Audible Behaviour at the Nest Box

Despite promising detection outcomes, the broader utility of using audio recorders as a remote monitoring tool is constrained by their feasibility. Factors such as classifier performance variability, the need for manual review, data storage demands, and the frequency of site visits for battery changes and data downloads must be carefully weighed when assessing feasibility.

The overall accuracy and specificity (i.e. non-target sounds were ignored) across all three classifiers would suggest they are performing well. However, a factor likely to skew accuracy and specificity performance metrics is the presence of class imbalance between target and non-target sounds. In this study, the true negatives far exceed the true positives (Table 2.7), inflating the classifiers' ability to correctly identify target sounds and avoid false detections. To accurately assess classifier performance, these metrics must therefore be interpreted alongside precision, sensitivity, and the Matthews Correlation Coefficient (MCC), which offer a more balanced evaluation in imbalanced datasets (van Merriënboer et al., 2024).

Across all classifiers, the skewed ratio of target to non-target events produced a well-established trade-off between sensitivity and precision - a pattern frequently observed in bioacoustic applications (Metcalf et al., 2022; Saito & Rehmsmeier, 2015). Increasing sensitivity improves the ability to detect rare target sounds but often comes at the cost of more false positives, reducing precision. Conversely, boosting precision tends to suppress false detections but risks missing true behavioural events. The ideal balance between these metrics depends on study goals. In this case, where the aim is to infer nest-level behaviour over time, higher sensitivity is preferable, even if it increases the need for manual review (Priyadarshani et al., 2018). Capturing a greater proportion of relevant vocalisations and nest box activity enables more comprehensive behaviour profiles of individuals (Teixeira et al., 2019).

Given the imbalance between target and non-target sounds, the classifiers' current performance in identifying target sounds already represents a substantial reduction in the manual workload. For example, Kaleidoscope processing reduced the total detections from tens of thousands in a week's worth of recordings to approximately 5,000, requiring an hour of manual review. Although feasible for a limited number of nest boxes in this study (n = 9), scaling up to generate a robust dataset will require identifying opportunities for methodological refinement and efficiency improvements.

2.4.6. Data Volume and Site Visit Frequency

Feasibility is also constrained by the data burden of continuous recording. Unlike camera traps, which often record only short clips upon trigger (e.g. ~1GB/week; Zárybnická et al., 2016), continuous audio recording with a 24kHz sample rate can generate up to 27GB per week. This creates a significant data storage demand, increasing the need for site visits to swap batteries and retrieve data; in this study's case, once per week, and increases the data processing time. A potential solution is to limit the recording schedule to periods of peak activity, such as three hours before and after sunrise. This targeted schedule could reduce data volume by ~75% and extend battery life, reducing the need for weekly site visits and data processing time. Previous work has shown that this approach can capture most relevant activity (Enari et al., 2019; Schackwitz et al., 2020; Bennett et al., 2025). Such alterations would not only reduce storage and processing requirements but also improve the practicality of deploying larger recording networks. However, the consequences of a reduced recording schedule on using detected behaviours to estimate breeding phenology and success must be tested.

2.4.7. Improving Classifier Performance

Further gains in reducing processing time will come from improving classifier performance, particularly by reducing false positives. The most effective way to achieve this is by expanding the training dataset to encompass a broader range of visitation and vocalisation types and acoustic conditions (Zhao et al., 2017).

Automated vocalisation detection software, such as BirdNET, had the greatest potential for improving time efficiency by reducing the need for manual review. Despite overall agreement between methods (Cohen's Kappa, Table 2.7), when the two approaches disagreed, custom-built classifiers in Kaleidoscope identified more vocalisations than the general models in BirdNET at two of the three nest boxes, as indicated by McNemar's test (Table 2.7). Although at nest box 12A, BirdNET produced more detections than Kaleidoscope indicating the need for further testing (Table 2.7). The localised models were likely better suited to capture vocalisations given the profile of the soundscape the in-situ recordings provided (Eberhardt, 2023). However, they also require more upfront effort and technical expertise.

A combined approach of limiting the recording schedule and refining site-specific classifiers is likely to offer the best pathway for scalable monitoring. Reducing the volume of incoming data lowers processing time, while more accurate classification reduces the need for manual review.

2.4.8. Toward Scalable and Interpretable Monitoring

Taken together, these results suggest that site-specific classifiers, even when developed from limited validated recordings, can still detect rare target vocalisations and visitations with sufficient sensitivity to support behavioural inference. While the classifiers may generate false positives, their value lies in applications where maximising detections is more important than precision, for example, when estimating visitation frequency or general activity patterns. Under this approach, low precision is managed through manual review, with classifiers serving to reduce the quantity of material to review, rather than as a tool capable of completing the full processing task on their own. Other acoustic behavioural monitoring projects have used a similar strategy, recognising that manual checks of positive detections are necessary and incorporating that requirement into their project timelines (Teixeira et al., 2019). The work required to correct for low precision will constrain the method's scalability. The extent to which manual review is feasible will depend on the mechanisms available to researchers to reduce the absolute number and proportion of false positives. For the methods proposed in this thesis, addressing the absolute number of false positives by limiting the recording schedule will reduce the total number of detections and manual verification effort. Equally, refining classifier performance with larger training datasets will likely improve precision by reducing the proportion of false positives. However, if further efficiencies need to be made, acoustic behavioural studies could learn from other applications of acoustic monitoring. Large-scale passive acoustic monitoring includes an additional filter by estimating the probability that a detection is correct and discarding those falling below a threshold (Funosas et al., 2024; Perez-Granados et al., 2025; Wood et al., 2023). Although typical applications of this method are used to estimate presence/absence, meaning the sensitivity of classifiers to capture all detections is not a priority. With each additional filtering process, the risk of diminished sensitivity also increases. Any gains in efficiency must be weighed against the risk of losing behavioural data. For example, using audio recorders at nest boxes to estimate breeding stages depends on detecting changes in nest visits over time, if processing efficiency is prioritised too heavily, there may come a point where missed detections prevent breeding phenology from being estimated accurately. Future efforts to streamline manual verification should evaluate the inclusion of additional filtering strategies to determine the point at which processing efficiency is maximized without compromising the detection of behavioural changes.

2.5 References

Araya-Salas, M., Grabarczyk, E. E., Quiroz-Oliva, M., García-Rodríguez, A., & Rico-Guevara, A. (2025). Quantifying degradation in animal acoustic signals with the R package baRulho. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.14481

Araya-Salas, M., & Smith-Vidaurre, G. (2017). warbleR: an r package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution*, *8*(2), 184–191. https://doi.org/10.1111/2041-210X.12624

Bennett, S., Williamson, L., Hernández-González, M., Denton, T., Laber, R., Deakin, Z., Bolton, M., Manilow, E., & Wilson, L. J. (2025). Bioacoustics as a Measure of Population Size and Breeding Success of European Storm Petrels Hydrobates pelagicus. *Ecology and Evolution*, *15*(8). https://doi.org/10.1002/ECE3.71893

Bejarano, V., & Jahn, A. E. (2018). Relationship between arrival timing and breeding success of intra-tropical migratory Fork-tailed Flycatchers (Tyrannus savana). *Journal of Field Ornithology*, *89*(2), 109–116. https://doi.org/10.1111/jofo.12251

Carlson, N. V., Pargeter, H. M., & Templeton, C. N. (2017). Sparrowhawk movement, calling, and presence of dead conspecifics differentially impact blue tit (Cyanistes caeruleus) vocal and behavioral mobbing responses. *Behavioral Ecology and Sociobiology*, *71*(9), 1–11. https://doi.org/10.1007/S00265-017-2361-X/FIGURES/2

Eberhardt, L. M. J. (2023). *Using remote surveying technology and acoustic analysis tools to infer the impact of infrastructure expansion on bird biodiversity in the tropical rainforests of Brunei, Borneo*. Durham University.

Enari, H., Enari, H. S., Okuda, K., Maruyama, T., & Okuda, K. N. (2019). An evaluation of the efficiency of passive acoustic monitoring in detecting deer and primates in comparison with camera traps. *Ecological Indicators*,*98*,753-762.https://doi.org/10.1016/J.ECOLIND.2018.11.062

Funosas, D., Barbaro, L., Schillé, L., Elger, A., Castagneyrol, B. and Cauchoix, M., 2024. Assessing the potential of BirdNET to infer European bird communities from large-scale ecoacoustic data. *Ecological Indicators*, *164*, p.112146. https://doi.org/10.1016/j.ecolind.2024.112146

Gaudon, J. M., McTavish, M. J., Hamberg, J., Cray, H. A., & Murphy, S. D. (2022). Noise attenuation varies by interactions of land cover and season in an urban/peri-urban landscape. *Urban Ecosystems*, *25*(3), 811–818. https://doi.org/10.1007/S11252-021-01194-4/FIGURES/3

Harrison, N. D., & Kelly, E. L. (2022). Affordable RFID loggers for monitoring animal movement, activity, and behaviour. *PLOS ONE*, *17*(10), e0276388. https://doi.org/10.1371/JOURNAL.PONE.0276388

Hereward, H. F. R., Facey, R. J., Sargent, A. J., Roda, S., Couldwell, M. L., Renshaw, E. L., Shaw, K. H., Devlin, J. J., Long, S. E., Porter, B. J., Henderson, J. M., Emmett, C. L., Astbury, L., Maggs, L., Rands, S. A., & Thomas, R. J. (2021). Raspberry Pi nest cameras: An affordable tool for remote behavioral and conservation monitoring of bird nests. *Ecology and Evolution*, *11*(21), 14585–14597. https://doi.org/10.1002/ece3.8127

Iverson, A. R., Schaefer, J. L. B., Skalos, S. M., Hawkins, C. E., Iverson, A. R., Schaefer, J. L. B., Skalos, S. M., & Hawkins, C. E. (2023). Global positioning system (GPS) and platform transmitter terminal (PTT) tags reveal fine-scale migratory movements of small birds: A review highlights further opportunities for hypothesis-driven research.

Ornithological Applications, *125*(3), 1–16. https://doi.org/10.1093/ORNITHAPP/DUAD014

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

Knight, E., Rhinehart, T., de Zwaan, D. R., Weldy, M. J., Cartwright, M., Hawley, S. H., Larkin, J. L., Lesmeister, D., Bayne, E., & Kitzes, J. (2024). Individual identification in acoustic recordings. *Trends in Ecology and Evolution*, *39*(10), 947–960.

Koh, J., & Opitz, T. (2025). Extreme-value modelling of migratory bird arrival dates: insights from citizen science data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *188*(3), 674–699. https://doi.org/https://doi.org/10.1093/jrsssa/qnae108

Lampe, H. M., Larsen, O. N., Pedersen, S. B., & Dabelsteen, T. (2007). Song degradation in the hole-nesting pied flycatcher Ficedula hypoleuca: Implications for polyterritorial behaviour in contrasting habitat-types. *Behaviour*, *144*(10), 1161–1178. https://doi.org/10.1163/156853907781890887

Marsland, S. (2011). Machine Learning : An Algorithmic Perspective. *Machine Learning*. https://doi.org/10.1201/9781420067194

Massimino, D., Harris, S. J., & Gillings, S. (2021). Phenological mismatch between breeding birds and their surveyors and implications for estimating population trends. *Journal of Ornithology*, *162*(1), 143–154. https://doi.org/10.1007/S10336-020-01821-5/FIGURES/6

Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote*

Sensing, *105*, 155–168. https://doi.org/10.1016/J.ISPRSJPRS.2015.03.014

Messmer, D. J., Alisauskas, R. T., Pöysä, H., Runko, P., & Clark, R. G. (2021). Plasticity in timing of avian breeding in response to spring temperature differs between early and late nesting species. *Scientific Reports*, *11*(1), 5410.

Metcalf, O. C., Barlow, J., Bas, Y., Berenguer, E., Devenish, C., França, F., Marsden, S., Smith, C., & Lees, A. C. (2022). Detecting and reducing heterogeneity of error in acoustic classification. *Methods in Ecology and Evolution*, *13*(11), 2559–2571. https://doi.org/10.1111/2041-210X.13967

Newson, S. E., Moran, N. J., Musgrove, A. J., Pearce-Higgins, J. W., Gillings, S., Atkinson, P. W., Miller, R., Grantham, M. J., & Baillie, S. R. (2016). Long-term changes in the migration phenology of UK breeding birds detected by large-scale citizen science recording schemes. *Ibis*, *158*(3), 481–495. https://doi.org/10.1111/IBI.12367

Pembury Smith, M. Q. R., & Ruxton, G. D. (2020). Effective use of the McNemar test. *Behavioral Ecology and Sociobiology*, *74*(11), 1–9. https://doi.org/10.1007/S00265-020-02916-Y/TABLES/5

Pérez-Granados, C., Funosas, D., Morant, J., Marín Gómez, O.H., Mendoza, I., Mohedano-Munoz, M.A., Santamaría, E., Bastianelli, G., Márquez-Rodríguez, A., Budka, M. and Bota, G., 2025. Optimization of passive acoustic bird surveys: a global assessment of BirdNET settings. *Ibis*. https://doi.org/10.1111/ibi.70013

Pérez-Granados, C. (2023). BirdNET: applications, performance, pitfalls and future opportunities. *Ibis*, *165*(3), 1068–1075. https://doi.org/10.1111/ibi.13193

Pérez-Granados, C., & Schuchmann, K.-L. (2020). Monitoring the annual vocal activity of two enigmatic nocturnal Neotropical birds: the

Common Potoo (Nyctibius griseus) and the Great Potoo (Nyctibius grandis). *Journal of Ornithology, 161*, 1129–1141. https://doi.org/10.1007/s10336-020-01795-4

Porlier, M., Charmantier, A., Bourgault, P., Perret, P., Blondel, J., & Garant, D. (2012). Variation in phenotypic plasticity and selection patterns in blue tit breeding time: Between- and within-population comparisons. *Journal of Animal Ecology*, *81*(5),1041–1051. https://doi.org/10.1111/J.1365-2656.2012.01996.X;SUBPAGE:STRING:FULL

Prinz, A. C. B., Taank, V. K., Voegeli, V., & Walters, E. L. (2016). A novel nest-monitoring camera system using a Raspberry Pi micro-computer. *Journal of Field Ornithology*, *87*(4), 427–435. https://doi.org/10.1111/jofo.12182

Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology*, *49*(5), jav-01447.

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/JOURNAL.PONE.0118432

Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., Woodruff, J., Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., & Woodruff, J. (2020). *Journal of Fish and Wildlife Management | www.fwspubs.org*. *11*. https://doi.org/10.3996/102019-JFWM-083

Stowell, D., Petrusková, T., Šálek, M., & Linhart, P. (2019). Automatic acoustic identification of individuals in multiple species: Improving identification across recording conditions. *Journal of the Royal Society Interface*, *16*(153).

https://doi.org/10.1098/RSIF.2018.0940;JOURNAL:JOURNAL:RSIF;REQUESTEDJOURNAL:JOURNAL:RSIF;PAGE:STRING:ARTICLE/CHAPTER

Surmacki, A., & Podkowa, P. (2022). The use of trail cameras to monitor species inhabiting artificial nest boxes. *Ecology and Evolution*, *12*(2), e8550. https://doi.org/10.1002/ECE3.8550

Teixeira, D., Maron, M., & van Rensburg, B. J. (2019). Bioacoustic monitoring of animal vocal behavior for conservation. *Conservation Science and Practice*, *1*(8), e72. https://doi.org/10.1111/CSP2.72

Trolliet, F., Huynen, M.-C., Vermeulen, C., & Hambuckers, A. (2014). B A Use of camera traps for wildlife studies. A review. *Biotechnol. Agron. Soc. Environ*, *18*(3), 446–454.

van Merriënboer, B., Hamer, J., Dumoulin, V., Triantafillou, E., & Denton, T. (2024). Birds, bats and beyond: evaluating generalization in bioacoustics models. *Frontiers in Bird Science*, *3*. https://doi.org/10.3389/fbirs.2024.1369756

Verhulst, S., & Nilsson, J. Å. (2007). The timing of birds' breeding seasons: a review of experiments that manipulated timing of breeding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1490), 399–410. https://doi.org/10.1098/RSTB.2007.2146

Walker, R. H., Robinson, R. A., Barimore, C. J., Blackburn, J. R., Barber, L. J., Clark, J. A., Clewley, G. D., Evans, H. E., Green, R. M., de Palacio, D. X., Grantham, M. J., Griffin, B. M., Schäfer, S., & Leech, D. I. (2018). Bird ringing and nest recording in Britain and Ireland in 2017. *Ringing and Migration*, *33*(2), 99–145. https://doi.org/10.1080/03078698.2018.1737454;JOURNAL:JOURNAL:TRAM20;PAGE:STRING:ARTICLE/CHAPTER

Wang, J. M., & Beissinger, S. R. (2011). Partial Incubation in Birds: Its Occurrence, Function, and Quantification. *The Auk*, *128*(3), 454–466. https://doi.org/10.1525/AUK.2011.10208

Wildlife Acoustics. (2024). *Kaleidoscope User Guide*. https://www.wildlifeacoustics.com/uploads/user-guides/Kaleidoscope11112024.pdf

Winiarska, D., Szymański, P., & Osiejuk, T. S. (2024). Detection ranges of forest bird vocalisations: guidelines for passive acoustic monitoring. *Scientific Reports*, *14*(1), 1–13. https://doi.org/10.1038/S41598-024-51297-Z;SUBJMETA=158,601,631;KWRD=ECOLOGY,ZOOLOGY

Wood, C.M., Barceinas Cruz, A. and Kahl, S., 2023. Pairing a user-friendly machine-learning animal sound detector with passive acoustic surveys for occupancy modeling of an endangered primate. *American Journal of Primatology*, *85*(8), p.e23507.

Zárybnická, M., Kubizňák, P., Šindelář, J., & Hlaváč, V. (2016). Smart nest box: A tool and methodology for monitoring of cavity-dwelling animals. *Methods in Ecology and Evolution*, *7*(4), 483–492. https://doi.org/10.1111/2041-210X.12509

Zhao, Z., Zhang, S. hua, Xu, Z. yong, Bellisario, K., Dai, N. hua, Omrani, H., & Pijanowski, B. C. (2017). Automated bird acoustic event detection and robust species classification. *Ecological Informatics*, *39*, 99–108. https://doi.org/10.1016/J.ECOINF.2017.04.003

2.6 Appendix – A

2.6.1 *F. hypoleuca* Nest Box Distribution Between 2018 to 2024

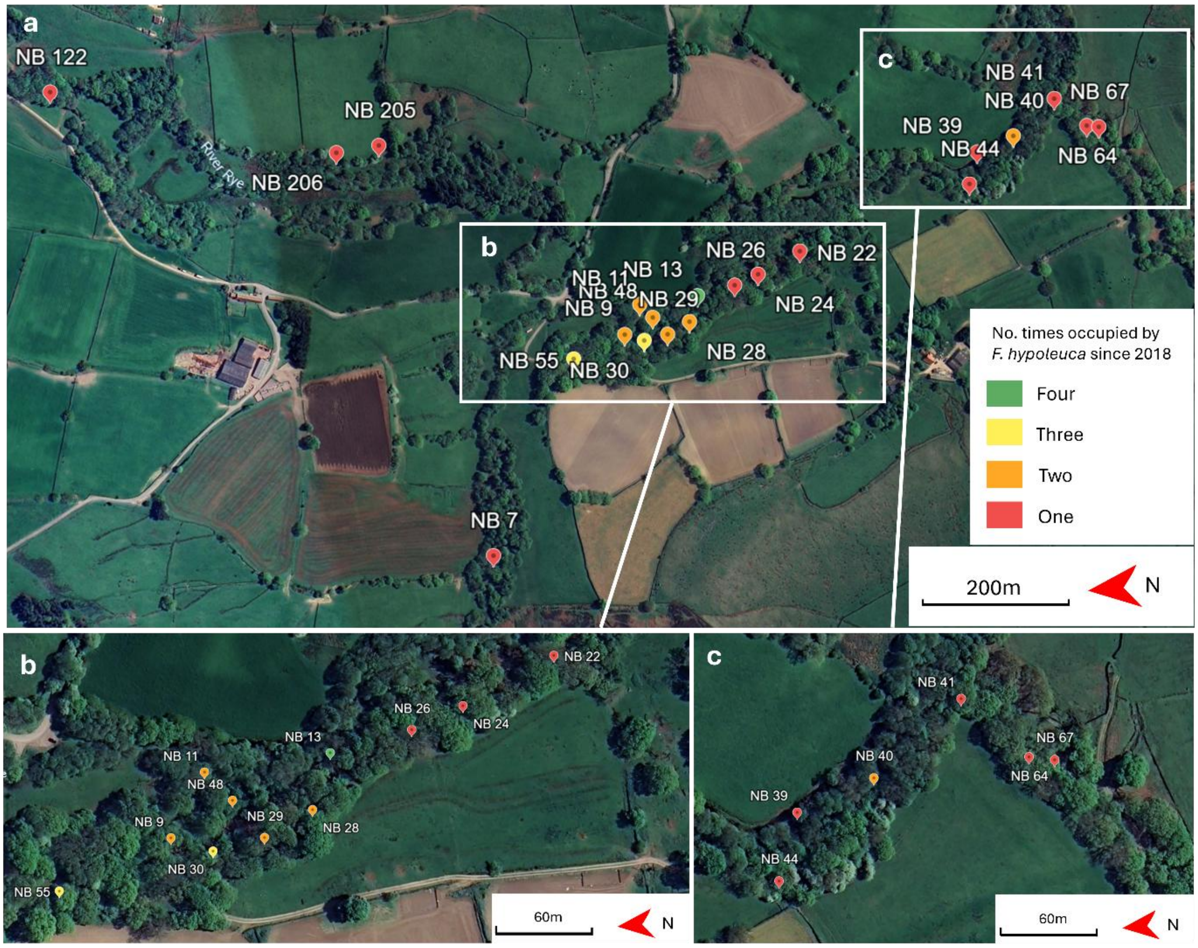

Figure 2.A1 Map of study site showing nest box occupancy history by Ficedula hypoleuca since 2018. (a) Overview of all monitored nest boxes, with insets (b) and (c) showing enlarged views of focal clusters. Colours indicate the number of years each nest box has been occupied: green = four, yellow = three, orange = two, and red = one. Base map imagery from Google Earth 2025.

2.6.2 Recorder Distribution in Relation to Previously Occupied *F. hypoleuca* Nest Boxes Between 2018 to 2024

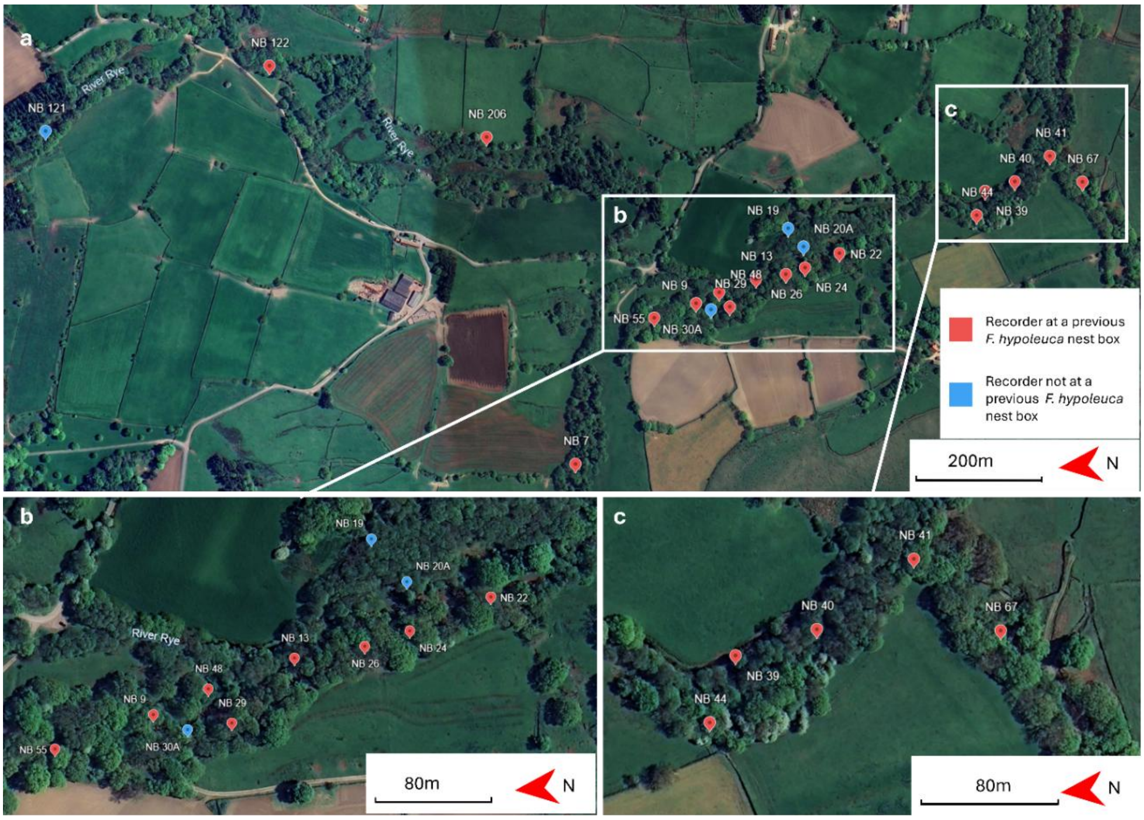

Figure 2.A2 Map of study site showing the nest boxes previously occupied by F. *hypoleuca between 2018 and 2024 with an audio recorder in red. Blue markers indicate nest boxes F. hypoleuca have not used between 2018 and 2024, recorders were installed to improve distribution and cover any potential blind spots if the target species moved to a previously unoccupied nest box. (a) Overview of all monitored nest boxes, with insets (b) and (c) showing enlarged views of focal clusters. Base map imagery from Google Earth 2025.*

2.6.3. Study Site Nest Box Occupancy Between 2018 and 2025

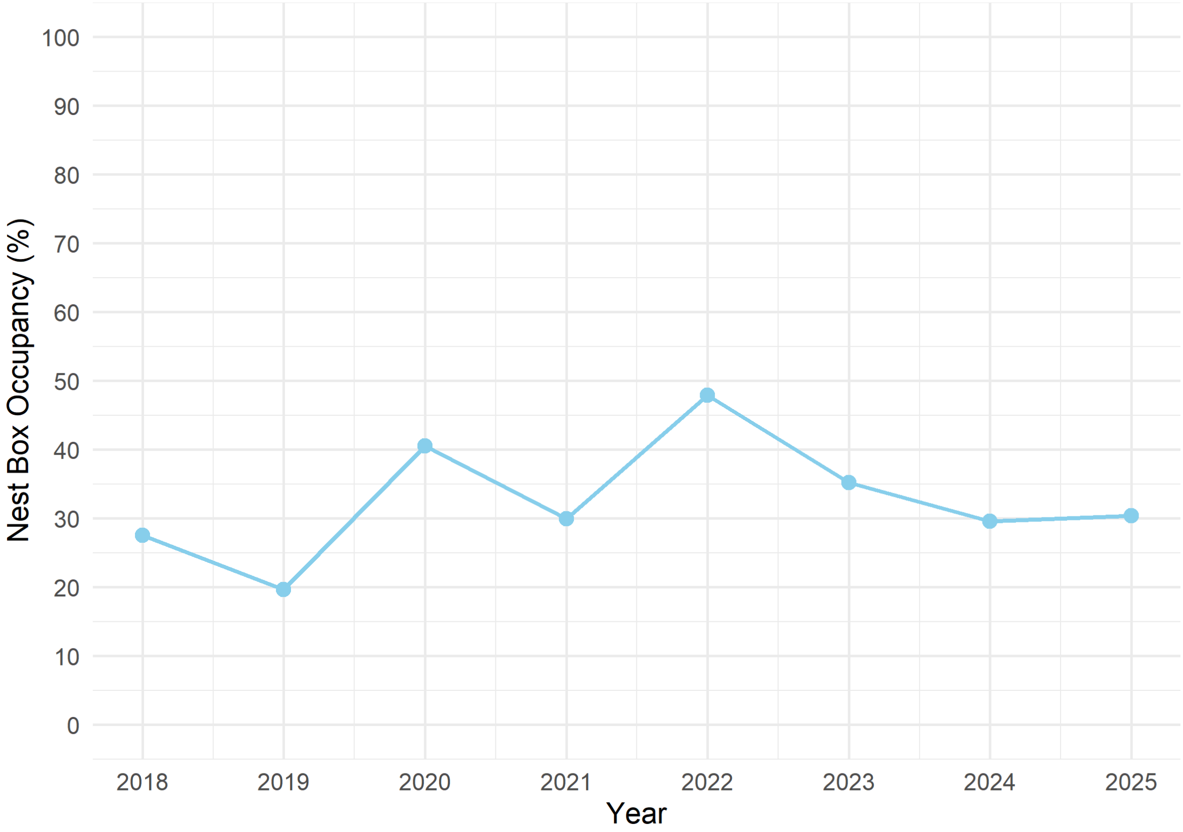

Figure 2.A3, Proportion of nest boxes occupied between 2018 and 2025. Overall occupancy varied considerably across years, ranging from 19% in 2019 to 48% in 2022. Occupancy generally remained between 30–40% for much of the study period, with no evidence of consistent directional change.

2.6.4. Study Site Nest Box Occupancy Between 2018 and 2025 Across Species

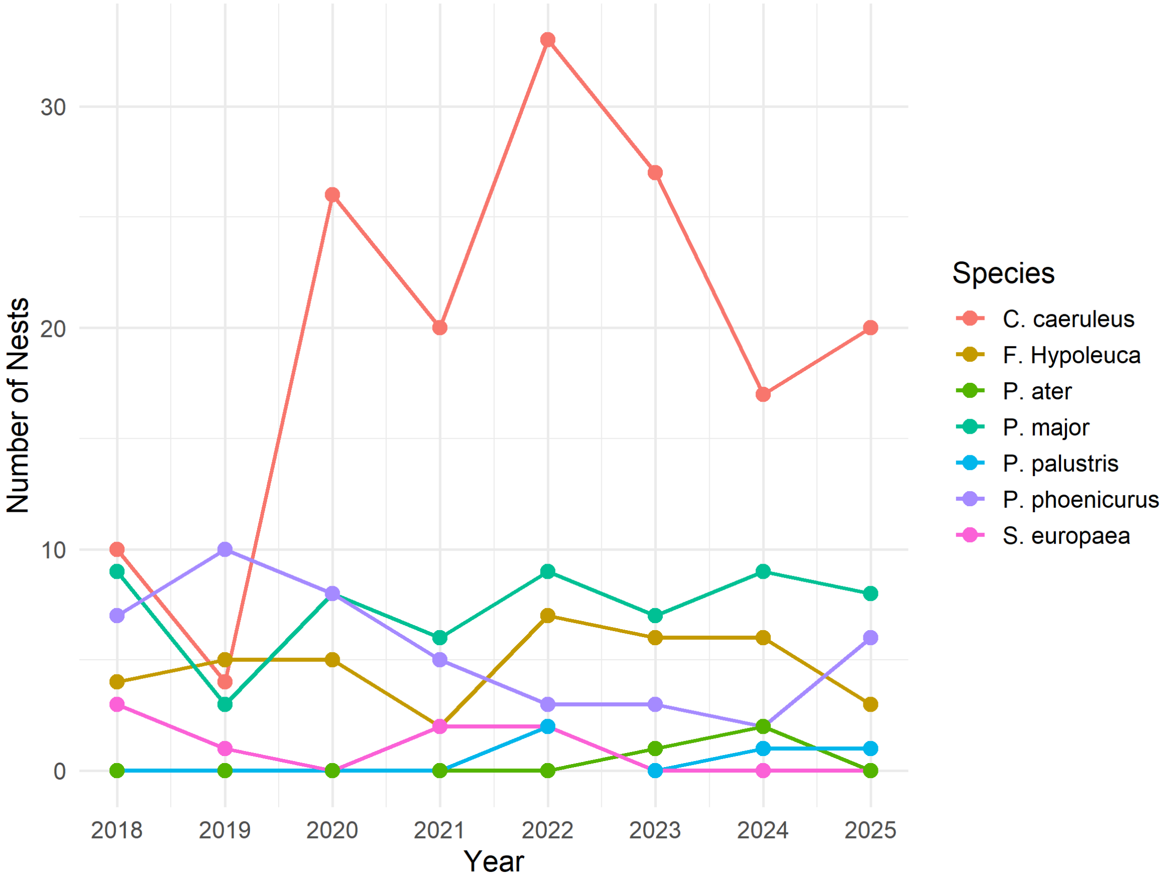

Figure 2.A4, Annual number of occupied nest boxes by species between 2018 and 2025. Species recorded include blue tit (Cyanistes caeruleus), pied flycatcher (Ficedula hypoleuca), great tit (Parus major), coal tit (Periparus ater), marsh tit (Poecile palustris), common redstart (Phoenicurus phoenicurus), and nuthatch (Sitta europaea). The focal study species of this study are C. caeruleus, P. major, and F. hypoleuca. C. caeruleus consistently occupied the largest number of boxes, peaking above 30 in 2022. P. major showed moderate but steady occupancy (typically six to nine boxes per year). F. hypoleuca occupied fewer boxes overall, with year-to-year variability ranging from two to seven boxes.

2.6.4. *F. hypoleuca* Productivity at Study Site Between 2018 and 2025

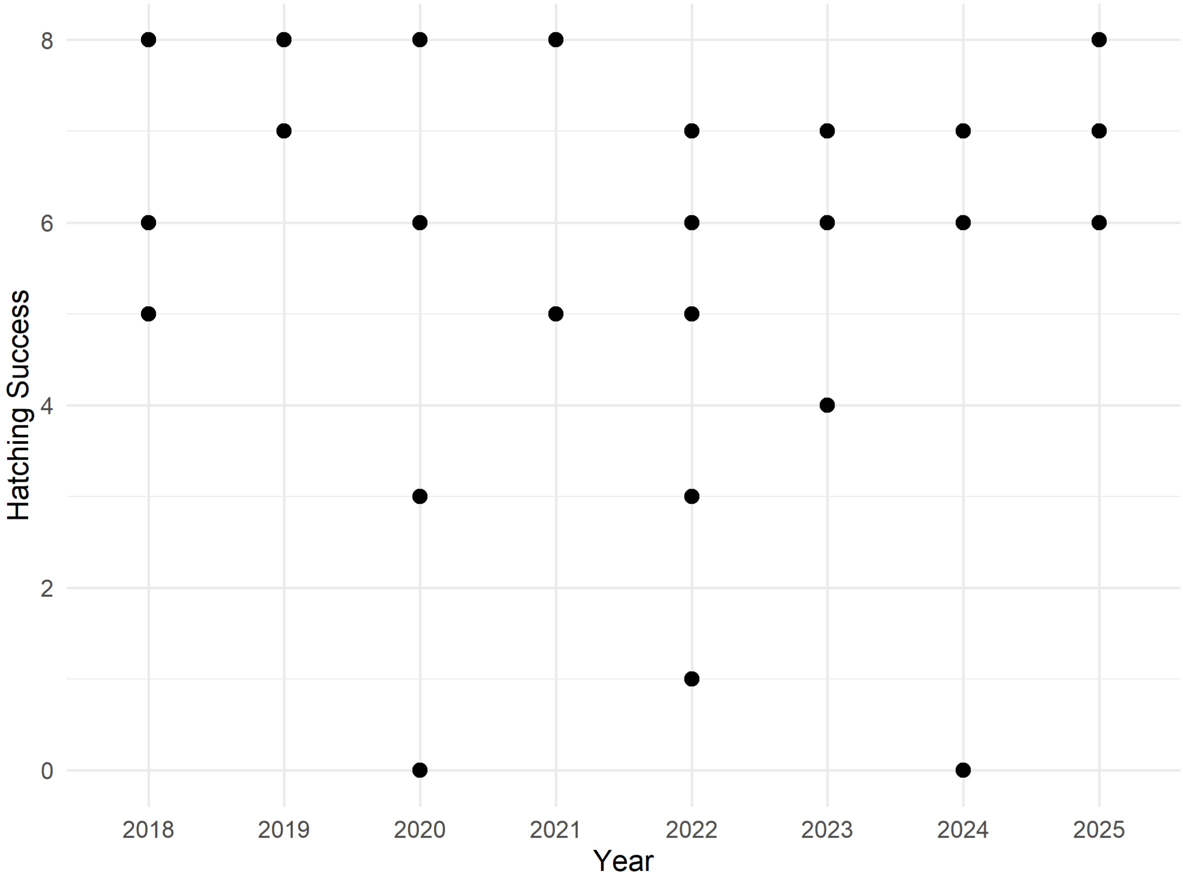

Figure 2.A5, Hatching success of F. hypoleuca nests between 2018 and 2025. Each point represents the number of chicks hatched in an individual nest. Most nests achieved high hatching success (6–8 chicks), but variation among years is evident, with some nests failing completely (0 chicks in 2020 and 2024) or producing fewer than three chicks. This highlights variability in breeding outcomes across years, despite generally high success in the majority of nests

2.6.5. Detection Range Assessment

Table 2.A1, Results of post-hoc Tukey HSD test comparing mean signal-to-noise ratio (SNR) across distances from the recorder in meters (m) and the tested frequencies in kilohertz (KHz). Confidence intervals are 95%. Comparison labels indicate the two distance intervals and frequencies being compared.

Comparison	Mean Difference in SNR Coefficients	95% CI Lower	95% CI Upper	Adjusted P-Value
Distance from Recorder				
1 - (-5)	21.978	14.874	29.083	1.29E-13
10 - (-5)	6.237	-0.766	13.239	0.118307
15 - (-5)	1.935	-5.170	9.039	0.990194
20 - (-5)	0.624	-6.379	7.627	0.999994
25 - (-5)	-4.259	-11.364	2.845	0.585431
5 - (-5)	15.353	7.697	23.009	2.98E-07
Background- (-5)	-3.050	-12.598	6.499	0.975332
10 - 1	-15.742	-22.627	-8.856	4.66E-09
15 - 1	-20.043	-27.032	-13.054	6.33E-13
20 - 1	-21.354	-28.240	-14.469	1.26E-13
25 - 1	-26.237	-33.226	-19.248	6.12E-14
5 - 1	-6.625	-14.174	0.924	0.12995
Background - 1	-25.028	-34.491	-15.565	1.81E-11
15 - 10	-4.302	-11.187	2.584	0.532658
20 - 10	-5.612	-12.393	1.168	0.182803
25 - 10	-10.496	-17.381	-3.610	0.000195
5 - 10	9.117	1.664	16.570	0.006063
Background - 10	-9.286	-18.673	0.101	0.054715
20 - 15	-1.311	-8.196	5.575	0.998953
25 - 15	-6.194	-13.183	0.795	0.122086
5 - 15	13.419	5.870	20.968	7.26E-06
Background-15	-4.984	-14.447	4.479	0.732115
25 - 20	-4.883	-11.769	2.002	0.364466
5 - 20	14.729	7.276	22.183	4.53E-07
Background-20	-3.674	-13.061	5.713	0.927326
5 - 25	19.613	12.064	27.162	3.76E-11
Background - 25	1.210	-8.253	10.673	0.999926
Background -5	-18.403	-28.287	-8.519	2.28E-06
Frequency				
4.5 – 3.8	2.045	-1.837	5.926	0.518
6.2 – 3.8	0.641	-3.295	4.577	0.974
No Tone – 3.8	0.899	-6.447	8.247	0.988
6.2 – 4.5	-1.404	-5.34	2.532	0.788
No Tone – 4.5	-1.145	-8.492	6.202	0.977
No Tone – 6.2	0.259	-7.117	7.635	0.999

3. Estimating Breeding Stage Phenology and Success Using Audible Behaviour

Abstract

Breeding phenology and success are key metrics for understanding how population dynamics respond to a changing climate. Traditional monitoring methods for these metrics are often labour-intensive, expensive, or intrusive. Increasingly, audio recording is recognised as a tool for detecting behaviours such as vocalisations and nest visitations. Detection of nest box behaviour could estimate breeding phenology and success metrics if the frequency of audible behaviour shifts in response to changing demands on parents across breeding stages and brood sizes.

In this study, Bayesian changepoints models were applied to counts of detected behaviour over time to test if changes in the frequency of these behaviours aligned with breeding phenology and metrics of breeding success. Compared with in-field estimates, visitation-derived changepoints were aligned with the onset of chick-begging and fledging, while vocally derived changepoints were aligned with estimates for arrival date at the breeding ground. Behaviour-derived estimates were consistent across both constant (24-hour) recording and reduced (continuous 6-hour) schedules. No relationship was found between visitation rates and brood size, although binary metrics of breeding success may be inferred.

These findings show that audio recorders can reliably estimate breeding phenology, including arrival date, but currently provide limited insight into breeding success. This highlights the potential of scalable acoustic monitoring to complement existing approaches for studying cavity-nesting birds.

3.1 Introduction

Uncertainty in ecological systems often leads individuals to adjust their behaviour to new conditions. The extent to which individuals can respond to change may profoundly affect both population dynamics and ecosystem functioning (Berger-Tal et al., 2011; Wong & Candolin, 2015). One domain in which climate and land-use change are driving behavioural changes is in breeding behaviour and phenology (Kellermann & van Riper, 2015; Newson et al., 2016; Nicolau et al., 2021). For migrant birds, the degree to which they can shift the timing of their breeding-ground arrival, egg-laying, or the fledging of young in response to seasonal changes can alter access to resources, influencing reproductive productivity and, consequently, long-term population viability (Kristensen et al., 2018; Romano et al., 2023; Shutt et al., 2022).

Typically, temporal changes in breeding behaviour and success have been observed using well-established methods of in-person nest box monitoring of cavity-nesting birds (Walker et al., 2018) commonly utilising citizen science networks (Nicolau et al., 2021). However, collecting comprehensive behavioural, phenological and reproductive success data is time-intensive and limited to trained individuals (Walker et al., 2018). The need for time- and cost-efficient monitoring capable of collecting comprehensives datasets on breeding phenology and success has, in some cases, driven the adoption of remote-sensing techniques, such as nest cameras which can provide information on visitation rates, hatching and fledging dates, and brood sizes (Hereward et al., 2021; Surmacki & Podkowa, 2022).

Increasingly acoustic monitoring is being recognised as an additional tool to infer behavioural shifts in response to their environment (Oestreich et al., 2024), and ultimately providing insight into conservation management (Teixeira et al., 2019). For instance, changes in song complexity have been linked to habitat quality in the migratory Bluethroat (*Luscinia svecica*) (García et al., 2024), providing an acoustic indicator for poor habitat quality and conservation prioritisation. In the Eurasian Eagle Owl (*Bubo bubo*), the timing of the dusk chorus reflects both habitat quality and reproductive fitness (Penteriani et al., 2014), information that can be used to identify high-quality breeding areas. More broadly, acoustic monitoring has been applied to assess habitat use under increasing anthropogenic disturbance (Ducrettet et al., 2020) and as a metric for breeding status and territory establishment (Digby et al., 2014; Upham-Mills et al., 2020). Such applications demonstrate how vocal behaviour can serve as a proxy for breeding success, allowing conservation managers to monitor species and habitats with minimal disturbance.

Despite the recognised potential of behavioural acoustic monitoring, little work has directly connected audio recordings to behaviour at each individual nesting site. Where acoustic approaches have been applied, they have focused on population-level breeding phenology (Schackwitz et al., 2020), and breeding success (Bennett et al., 2025) or at the nest-level, detecting occupancy and whether breeding attempts were successful or not (Teixeira et al., 2022). A logical extension of this work is to test the potential for acoustic monitoring within long-standing study systems of cavity-nesting species, where nest boxes provide reliable focal points for detailed observations of behaviour, phenology, and reproductive success, enabling estimations of population-level dynamics and responses to environmental change (Walker et al., 2018). In this context, the approach developed in Chapter 2 can be used to evaluate whether remotely detected audible behaviours can be used to estimate breeding behaviour, phenology, and success. Such estimations could rely on the simple presence or absence of audible behaviours, or on the frequency with which those behaviours are detected over time.

There are many changing demands on nesting individuals across the breeding period, from territory establishment and mate-pairing through vocalisations (Nowicki & Searcy, 2005), increased nest box visitation for food provisioning post-hatching (Ritchison, 2023), to the absence of visitations post-fledging. It is reasonable to conjecture that shifts in these demands correspond to detectable changes in audible cues in the vicinity of the nest box. Specifically, transitions between breeding stages, such as the onset of incubation, hatching, and fledging, should be reflected in shifts in the audio data, as illustrated in Figure 3.1b. If so, acoustics could provide a remote monitoring tool capable of delivering phenological data comparable to traditional nest box monitoring (Walker et al., 2018). In addition, the onset of frequent vocalisations around or visits to the nest box could offer a means of more accurately estimating arrival dates, and estimating them remotely, thereby extending the nest box monitoring toolkit whilst minimising disruption (Bejarano & Jahn, 2018).

Estimating breeding stage transitions requires identifying the time at which the frequency of events such as vocalisations and visitations changes. Changepoint analysis provides a method to detect such shifts, often also referred to as a broken stick, broken line, or segmented regression analysis (Lindeløv, 2020). Although less commonly used in ecological studies, examples include estimating thresholds in environmental conditions that trigger changes in seedling recruitment (Beckage et al., 2007), or analysing telemetry data of turtles to detect shifts in diving events to distinguish between migrating or foraging behaviour (Patel et al., 2015). In the context of breeding birds, a changepoint analysis provides a robust framework to quantify the dates at which there is a sustained change in the frequency of vocalisations or visitations along the breeding timeline. These estimated dates can then be compared to the timing of breeding stage transitions (e.g. from incubation to post-hatching) to assess alignment.

To evaluate whether data collected from audio recorders can inform population dynamics and conservation management, the audible breeding behaviour must be a known proxy for measures of breeding success (Berger-Tal et al., 2011). In some passerines, visitation rates can have a strong positive relationship with brood size (Madden et al., 2022; Sanz & Tinbergen, 1999). Demonstrating this link with audio-derived visitation data, as illustrated in Figure 3.1c, would provide confidence that data from acoustic monitoring can serve as reliable proxies for indicators of breeding success such as brood size (Dunn & Winkler, 2010; Shutt et al., 2022) or chick weight gain (Moreno et al., 1997). Beyond method validation, the scalability of audio monitoring offers opportunities to examine how variation in visitation rates influences these outcomes concurrently across multiple sites.

If breeding phenology and success are detectable, feasibility would become the primary barrier to adopting the method at a larger scale. The 24-hour recording protocol described in chapter 2 requires weekly battery replacement and data retrieval. If estimates of breeding phenology, behaviour, and reproductive success depended strictly on constant recording, the method would offer little advantage over traditional nest-box checks, since the field effort would remain comparable (Walker et al., 2018). For the method to function as a truly scalable remote-monitoring tool, recording schedules would need to resemble existing large-scale passive acoustic monitoring practices, which concentrate effort during peak activity periods such as dawn and dusk to reduce the effort required for data storage and processing (Sugai et al., 2019). However, it is not yet clear whether shifts in visitation or vocalisation rates across breeding stages follow consistent daily patterns that a reduced recording schedule could reliably detect. For instance, if adults respond to heightened post-hatching provisioning demands by increasing visit frequency throughout the entire day rather than concentrating their activity at dawn and dusk, a reduced recording window would miss this change. To test this, the constant recording effort used in this study provides a benchmark against which the estimates derived from reduced effort strategies can be evaluated.

On this basis, this chapter applies the methodological framework established in Chapter 2 to detect behaviours at the nest box (Figure 3.1a), to evaluate the utility of acoustic monitoring for studying breeding behaviour, phenology and success.

Specifically, this chapter had two main objectives:

1. To determine whether changes in the frequency of audible behaviours can be used to estimate breeding stages, including individual arrival dates, onset of incubation, hatching, and fledging (Figure 3.1b)

2. To evaluate the relationship between detected behaviours and breeding outcomes, such as how parental visitation rates relate to chick growth and brood size (Figure 3.1c)

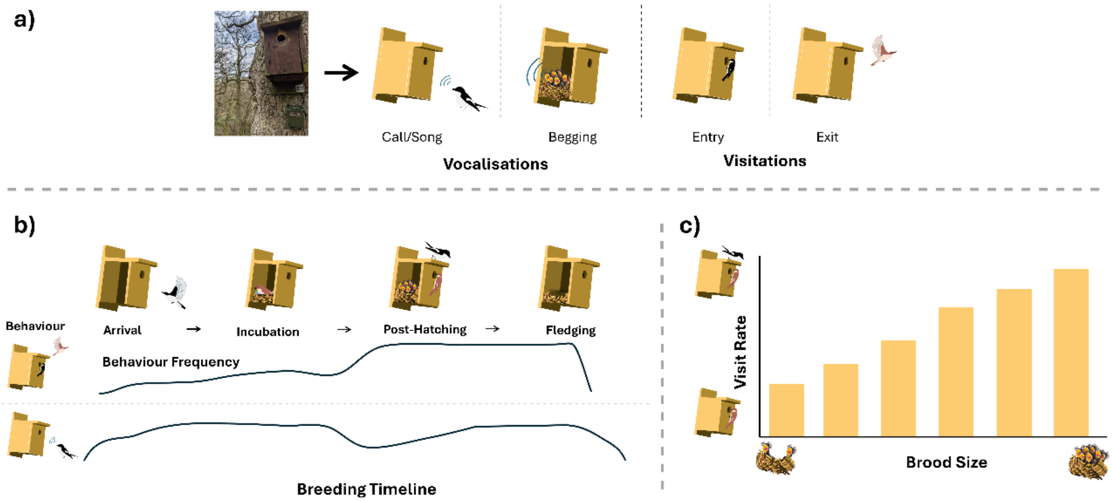

Figure 3.1, a) Schematic representing the forms of behaviour audio recorders are capable of detecting. b) Schematic representation of how behaviours such as activity (nest box entry/exit) and vocalisations may change across breeding stage transitions, between arrival and incubation, incubation and post-hatching, and post-hatching to fledging. c) Schematic representation of the potential relationship between brood size and the frequency of visitation (e.g. entry and exit events).

3.2. Methods

3.2.1. Nest Box Monitoring

The study site was the same as described in Chapter 2 (see also Figure 2.2 for recorded nest box distribution across the site).

The 20 nest boxes with recorders were monitored every 6-7 days from March 27th to 12th June 2025 by AD. To determine nest box occupancy across the site, all 125 nest boxes were monitored every 6-7 days from 15th April to 7th May 2025 by AD, R-C, and R-A. All occupied nests were monitored following standard protocols (Walker et al., 2018).

The primary researcher (AD) was present for all data collection at the study site and held a T-permit from the British Trust for Ornithology (BTO). In accordance with the Nest Recording Scheme Code of Conduct, monitoring of nest boxes does not require a permit from the BTO (British Trust for Ornithology, n.d.). Any monitoring required at the nest box, without handling outside the box, such as brood counts could be done independently by AD. Any handling of birds for measurement purposes was carried out under the supervision of two licensed recorders: R-C, who held a C-permit, and R-A, who held an A-permit, both issued by the BTO with unrestricted endorsements for ringing and handling cavity-nesting species.

Once eggs in the nest hatched, the nest checks increased to every three days by AD and R-C. Post-hatching, breeding success was measured as hatching success (number of eggs hatched) and fledging success (number of hatched chicks that fledged). For six of the nine recorded nests, chick weight gain of *P. major and C. caeruleus nests* were recorded by AD and R-A. Chicks were weighed at approximately day 4-5 post-hatching and again at day 11-12 to capture the period of greatest weight gain (Kern et al., 2001; Moreno et al., 1997). All chicks within a brood were weighed, and the mean value was used in analyses. All procedures were approved by the Durham University Animal Welfare Ethical Review Board and carried out under BTO licences.

In-field Estimates for Breeding Stage Transition

The in-field monitoring provided reference records for the approximate onset of a breeding stage transition from egg laying to incubation, incubation to hatching, and hatching to fledged (Figure 3.2). The date of the breeding stage transition estimated from in-field observations, could then be related to the audio data.

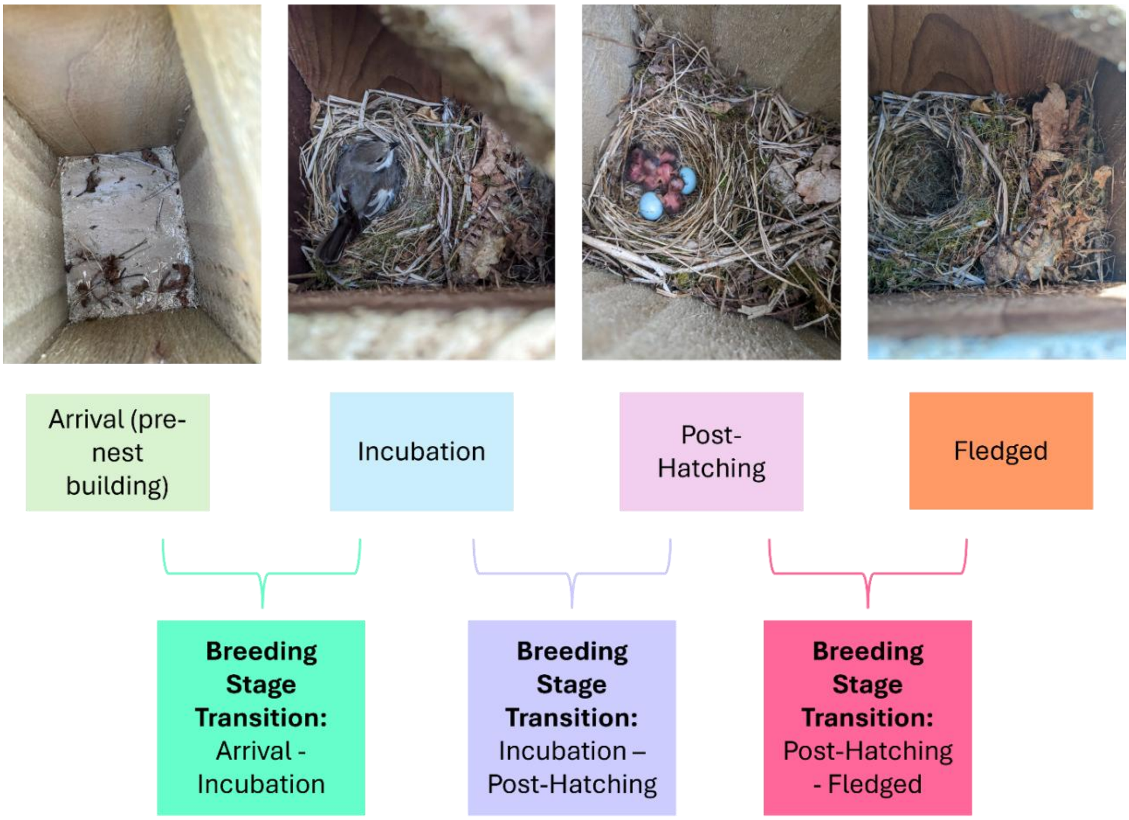

Figure 3.2 Breeding stages of F. hypoleuca within a nest box, showing arrival (pre-nest building), incubation, post-hatching, and fledged stages. Coloured boxes highlight each stage and the transition between each breeding stages. In-field estimates of the breeding stage transition date were related to the audio-derived changepoints.

Migrant arrival dates were inferred from vocalisations rather than directly monitoring visitation at the nest box, as recorders were not initially on the same trees that eventually held *F. hypoleuca* nest boxes (Chapter 2, Figure 2.3). The first vocalisation detected by a nest box recorder was used as the reference for an individual's estimated time of arrival at the vicinity of the nest box. Each estimated time of arrival to a nest box was then related to a vocally-derived changepoint. With only three *F. hypoleuca* nests, the analysis to estimate arrival dates was applied only to these three nests.

Available Audio Data for Changepoint Analysis of Breeding Stage Transitions

The availability of audio data across nest boxes, arrival dates, and breeding stages is summarised in Table 3.1. The transition between arrival and incubation for *F. hypoleuca* was not estimated using visitation detections due to the reduced recording period prior to onset of incubation (1-2 Days) as audio recorders were being moved during this stage (Chapter 2, Figure 2.3). Vocalisation and visitation rates were analysed independently, with no combined analysis.

Table 3.1 Available audio data for each nest box, summarised in relation to estimated arrival date and in-field estimates of the onset of breeding stages. For each stage, visit refers to adult entry and exit events, while vocal refers to songs and calls. Cells are colour coded: green = "y" (behaviour monitored at this stage); red = "n" (behaviour not monitored at this stage); grey = "N/A" (arrival date not applicable to resident species P. major and C. caeruleus and vocalisations of P. major and C. caeruleus were not monitored).

Nest Box	**Species**	**Arrival**		**Incubation**		**Hatched**		**Fledged**	
		Visit	Vocal	Visit	Vocal	Visit	Vocal	Visit	Vocal
12A	*F. hypoleuca*	n	y	n	y	y	y	y	y
14A	*F. hypoleuca*	n	y	n	y	y	y	y	y
37A	*F. hypoleuca*	n	y	n	y	y	y	y	y
13	*P. Major*	N/A	N/A	y	N/A	y	N/A	y	N/A
39	*P. Major*	N/A	N/A	y	N/A	y	N/A	y	N/A
20A	*C. caeruleus*	N/A	N/A	y	N/A	y	N/A	y	N/A
22	*C. caeruleus*	N/A	N/A	y	N/A	y	N/A	y	N/A
41	*C. caeruleus*	N/A	N/A	y	N/A	y	N/A	y	N/A
44	*C. caeruleus*	N/A	N/A	y	N/A	y	N/A	y	N/A

3.2.2. Using Audio-Derived Changepoints to Estimate Breeding Stage Transition

For this study, a changepoint is defined as the time at which the frequency of a detected behaviour shifts significantly, as estimated by a changepoint model. The behaviours analysed here are visitation and vocalisation events, considered separately. Visitation events were detected using Kaleidoscope Pro classifiers, which was the sole detection method for this behaviour applied in the thesis. Vocalisation detections were also from the Kaleidoscope Pro classifiers, as this method tended to detect more vocalisations compared to BirdNET (Chapter 2, Table 2.7; McNemar’s Test). Both visitation and vocalisation events were aggregated into hourly counts. Tests at daily and 15-minute resolutions indicated that daily aggregation masked within-day changes in behaviour patterns, while finer 15-minute aggregation introduced strong daily cycles that risked being misidentified as changepoints. Hourly aggregation therefore provided the most reliable resolution for changepoint analysis. Hours with zero detections were retained to preserve a regular time series.

The changepoint method used in this analysis was a Bayesian changepoint Poisson regression in the package mcp (Lindeløv, 2020), with the analysis conducted in R Studio (version 4.1.1.) (R Core Team 2024).

For audio data to estimate the timing of a breeding stage transition changepoint models were applied to subsets of data restricted to a single transition. For example, hourly visitation counts spanning incubation through post-hatching were used to estimate the changepoint between these two stages (Figure 3.3). This was done for each breeding stage in series, i.e. arrival to incubation, incubation to post-hatching, post-hatching to fledging across both behaviours and all nest boxes with the available data (Table 3.1).

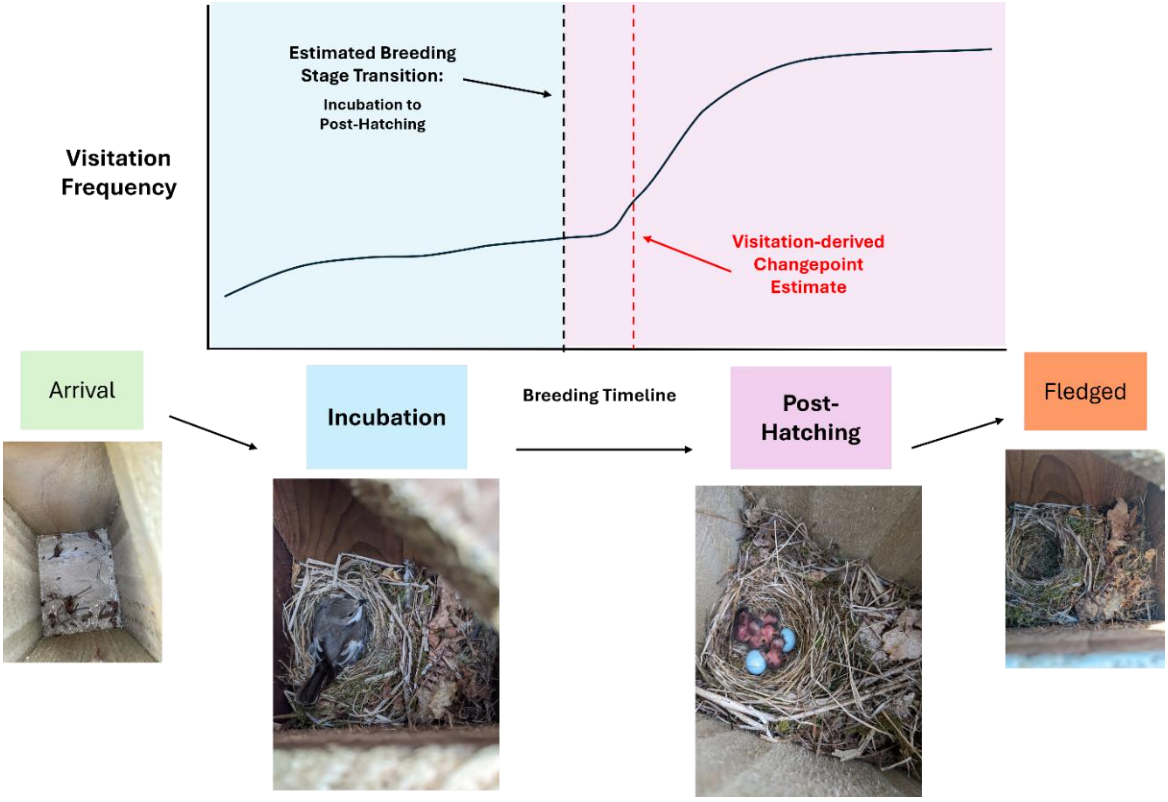

Figure 3.3, A schematic representing the data including two breeding stages used for each changepoint analysis, with a comparison between estimated breeding stage transition and a visitation-derived changepoint estimate. The beginning of the blue section is the onset of incubation, with the pink corresponding to the estimated post-hatching period, separated by the estimated breeding stage transition date. Incubation follows the estimated arrival, and post-hatching period preceding the fledged nest. Visitation frequency on the y-axis is one of two behaviours that are used to derive a changepoint estimate, with the second being vocalisation frequency.

In a changepoint analysis, the data are partitioned into segments separated by changepoints, where each segment represents a period with consistent statistical properties (e.g. a constant mean or linear trend). Unlike a frequentist framework, which treats changepoints and segment parameters as fixed but unknown quantities estimated solely from observed data, the Bayesian approach models them as random variables with prior distributions, allowing prior knowledge to be incorporated and uncertainty to be expressed directly through posterior distributions and credible intervals. These priors can encode expectations about the Prior distributions can incorporate expectations about the timing and structure of these segments, allowing existing knowledge to inform model estimation. For example, in Figure 3.4a, the onset of hatching is followed by a sudden increase in visitation frequency, after which visitation continues to rise gradually. This pattern is represented as a changepoint followed by a linear Julian day slope, such that the post-changepoint mean increase steadily with each successive day instead of remaining constant. Alternatively, Figure 3.4b is best described by a single changepoint separating two segments with distinct constant means. Posterior distributions are obtained via Markov Chain Monte Carlo sampling, allowing uncertainty in changepoint timing and segment trends to be quantified directly through credible intervals.

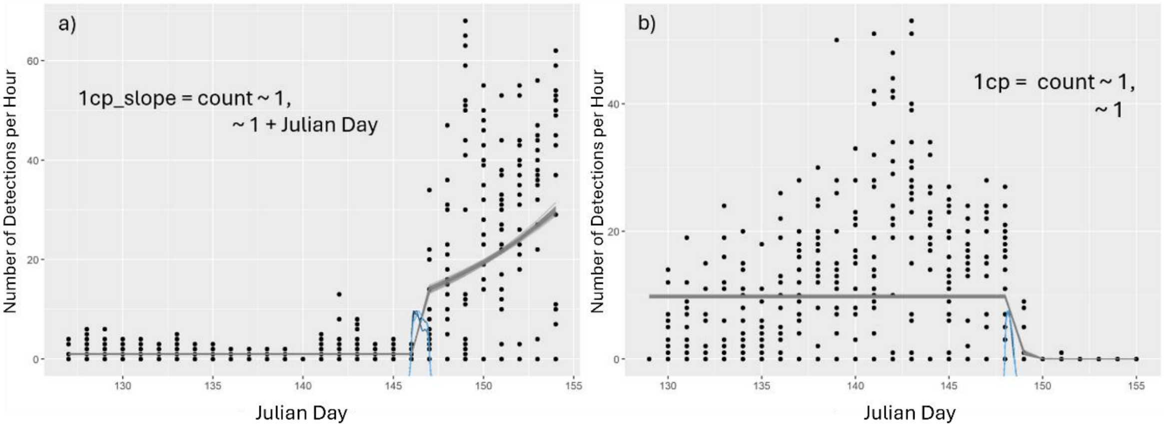

Figure 3.4 *Posterior model fits of the number of detections per hour across Julian days. Black points represent observed hourly counts. Black lines show the posterior mean fit. Blue lines indicate the posterior distribution of estimated change points (cp). a) Model with a single change point and a slope parameter (1cp_slope = count ~ 1 + Julian Day), representing the transition from incubation to post-hatching. b) Model with a single change point and intercept only (1cp = count ~ 1, ~1), representing the transition from post-hatching to fledging. All posterior model fits are provided in Appendix B, section 3.6.1 & 3.6.2.*

A set of candidate models were applied to the frequency of vocalisation and visitation data set out in Table 3.1. The candidate models, varied in (i) number of changepoints (0–3), (ii) whether the post-changepoint mean included a linear trend with Julian day (i.e. including a slope term to capture gradual increases or decreases over time), and (iii) presence/absence of an autoregressive term for serial dependence. The candidate models were chosen based on prior observation of the data and recommendations for building segment structures as described by Lindeløv (2020).

Model adequacy was evaluated using posterior predictive checks to assess whether simulated data resembled the observed data. Predictive performance was then compared across models using Pareto smoothed importance sampling leave-one-out cross validation (PSIS-LOO) as implemented in the loo package (Gabry, 2020) (PSIS-LOO). The reliability of these PSIS-LOO estimates was assessed using Pareto-K diagnostics, which quantify how influential each data point is in the cross-validation procedure. If leaving out a particular point would change the posterior distribution too strongly, the importance-sampling approximation becomes unstable, which is indicated by a high k value. Following Gabry (2020), data points with k values above 0.7 were treated as problematic. Models with fewer than three data points exceeding the 0.7 threshold were retained, since this level of discrepancy is still regarded as reliable (Vehtari et al., 2017). For each dataset, models failing Pareto-k diagnostics were excluded, and the best-fitting model based on PSIS-LOO among those retained was chosen. This model provided the changepoint between events in terms of median Julian day, along with the 95% credible interval.

In-field estimates of breeding stage transition dates were used as a reference to assess whether changepoints derived from audio data could detect these transitions. For each breeding stage transition, the difference between the audio-derived median date and its 95% credible interval from the changepoint analysis and the in-field estimated date was calculated. Across nest boxes, the deviations of the median and 95% credible intervals were reported for each breeding stage transition and behaviour - visitation and vocalisations.

3.2.3. Testing the Audio Effort Required to Detect Breeding Stages

Previous changepoint estimates in this chapter have used visitation and vocalisation detections from a constant (24 hour) recording effort. The audio effort required to detect a breeding stage transition was evaluated by applying the changepoint analysis (as described in 3.2.2. of this chapter) to audio data from three reduced recording schedules.

Visitation and vocalisation detections were filtered into three recording schedules: (1) three hours post-sunrise and prior to sunset, (2) six hours post-sunrise, and (3) six hours prior to sunset. These schedules were used based on the peak periods of daily visitation and vocalisation detections, detections were taken from the whole breeding period adjusting for changes in the timing of sunrise and sunset.

The extent to which the estimated changepoints from reduced recording schedules differed from those estimated using the 24-hour audio data was tested with a linear mixed-effects model (LMM). The difference in the estimated changepoint was the response variable, with recording schedule as a fixed effect, and nest box and breeding stage included as random effects. The LMM was run using the lmerTest (Kuznetsova et al., 2025), with the analysis conducted in R Studio (version 4.1.1.) (R Core Team 2024).

For a visual comparison of recording schedule performance, the median estimates and 95% credible intervals calculated during the changepoint analysis were averaged across the nine boxes to obtain mean estimates and intervals for each recording schedule and breeding stage transition.

3.2.4. Relating Visitation to Breeding Success

As the number of nests recorded was limited to nine, modelling the relationship between visitation and breeding success would be unreliable and so was not attempted in this study. However, to visualise parental visitation patterns derived from audio data across brood sizes and species, the number of visits per chick were plotted against total hatching and fledging number for each nest, separately for each species: *C. caeruleus, P. major, and F. hypoleuca*. This approach allows comparison of provisioning effort relative to brood size.

Chick weight gain was measured in only four broods, as nest box monitors were reluctant to take additional measurements of *F. hypoleuca* chicks due to a relatively smaller number of nests (Chapter 2, Figure 2.A4). Once chick weight gain metrics were agreed to be taken, a *C. caeruleus* and a *P. major* nest were at post-hatching age of approximately day 10, thereby missing early growth. To ensure chick handling only yielded valuable data, these nests were not measured. The relationship between chick weight gain and total visitation rate between measurement dates is shown in Appendix B, figure 3.B8. However, the small sample size (n = 4) did not allow for meaningful statistical analysis, and further data collection will be required to explore this relationship.

3.3. Results

3.3.1. Visitation-derived Changepoints

To illustrate how the changepoint models capture changes in visitation patterns for each nest box, the posterior fit of hourly detection counts across Julian day is provided in Appendix B Section 3.6.1. The plots show visitation patterns alongside posterior mean fits and the posterior distribution of estimated changepoints for each breeding stage transition.

The agreement between the visitation-derived changepoints and the in-field observed changepoint varied by breeding stage, as shown in Figure 3.5. The median deviation from the field-inferred changepoint was -2.4 days (IQR: -9.9 - 2.9) for incubation, +5.5 days (IQR: 0.8 - 6.5) for hatching, and -1.8 days (IQR: – 3.5 - -1.2) for fledging. Credible intervals at 95% followed a similar pattern: onset of hatching and fledging were associated with narrow intervals, with median widths of 0.95 days (IQR: 0.95 - 0.95) and 0.94 days (IQR: 0.93 - 0.95), respectively, while incubation intervals were substantially wider, with a median width of 3.46 days (IQR: 2.1 - 13.9).

Between species, the median deviation (days) from the in-field estimate of incubation onset was -0.5 (IQR: -6.6 – 5.6) in *P. major,* and -2.4 (IQR: -6.9 – 1.5) in *C. caeruleus*. The deviation in estimating hatching onset was -0.44 (IQR: -0.47 - -0.41) for *P. major*, +5.5 (IQR: 5.5 – 6.5) for *F. hypoleuca*, and +6.5 (IQR: 5.1 – 6.7) for *C. caeruleus.* For fledging onset, deviated estimates were -1.6 (IQR: -1.7 - -1.5) for *P. major*, -0.52 (IQR: -0.83 - -0.02) for *F. hypoleuca*, and -7.2 (IQR: -10.9 - -3.3) for *C. caeruleus*.

Across all three species, credible intervals were relatively narrow for estimating fledging onset (median width: 0.71, IQR: 0.60–0.83 in *P. major*; 0.95, IQR: 0.95–1.17 in *C. caeruleus*; and 0.93, IQR: 0.58–0.95 in *F. hypoleuca*). For hatching onset, intervals were also narrow in *C. caeruleus* (median width: 0.95, IQR: 0.95–1.92) and *F. hypoleuca* (0.95, IQR: 0.95–0.95), whereas *P. major* showed substantially wider intervals (3.46, IQR: 2.20–4.71). Estimates for the onset of incubation had the widest intervals overall in *C. caeruleus* (median width: 9.49, IQR: 1.67–18.01), while *P. major* showed narrower intervals (3.46, IQR: 3.05–3.86).

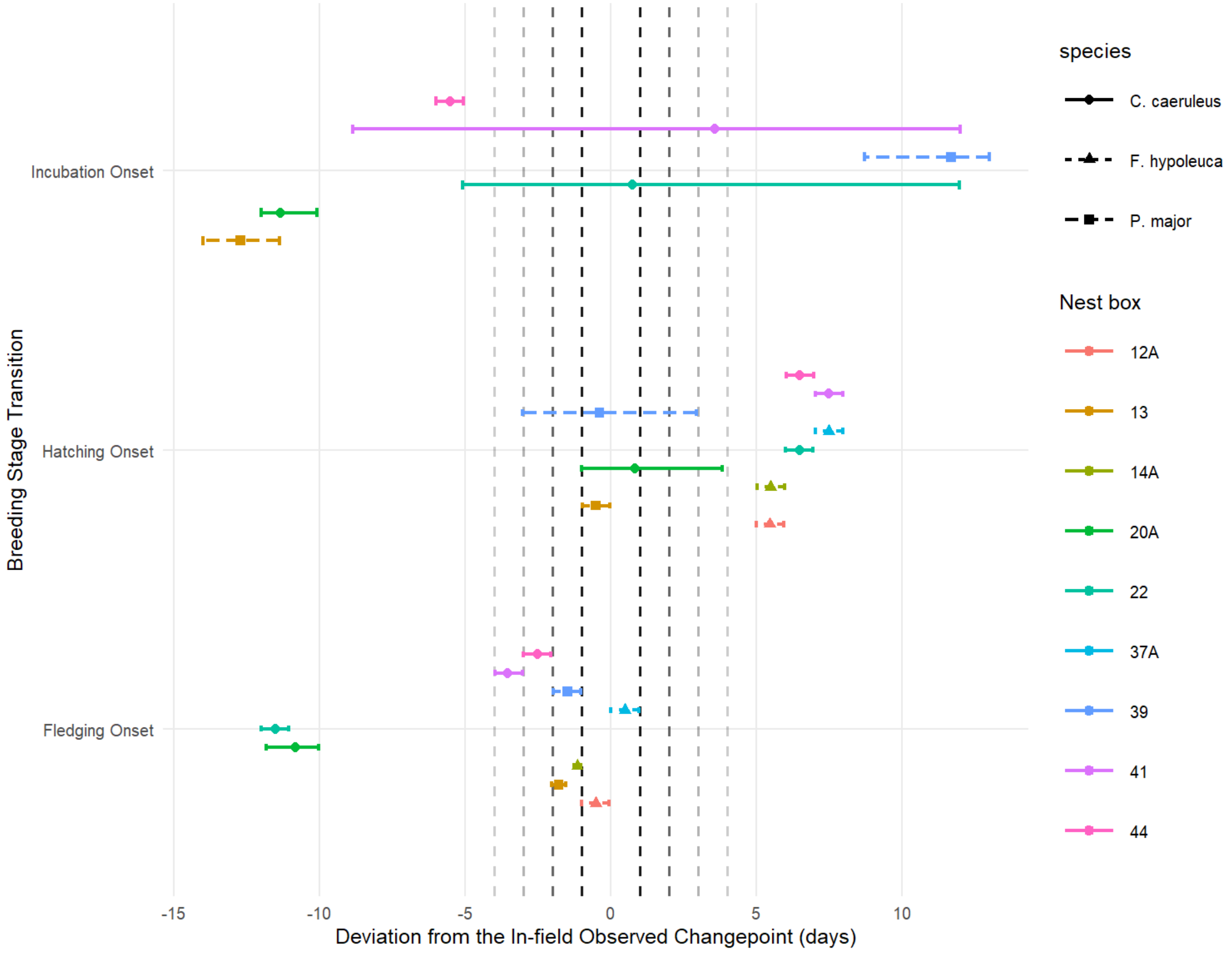

Figure 3.5 Changepoints, derived from models fitted to visitation detections from acoustic data across all nine studied nest boxes, compared with field-estimated breeding stage transitions. Points represent estimated changepoints for individual nest boxes, relative to the field-inferred date of the event. Horizontal lines showing 95% credible intervals from the changepoint model (Appendix B, Section 3.6.1.). Vertical dashed lines indicate the in-field inferred changepoints (0) and +/- 1–5 days around the observed date (grey). Colours correspond to individual nest boxes, see Chapter 2, Figure 2.2 & 2.3. Line type and shape correspond to the species. Breeding stage transitions are shown separately for incubation onset, hatching onset, and fledging onset.

To visualise the distribution of visitation detections along the breeding timeline, the mean detection count per hour for each nest box is plotted against Julian day relative to the in-field estimate for hatching (Figure 3.6). Detections remained relatively low during nest building and egg laying, with a slight increase during incubation although not sufficient for the changepoint analysis to reliably estimate onset of incubation (Figure 3.5). For the onset of hatching, the changepoint analysis estimate was consistently delayed by approximately 5 days (Figure 3.5), the extent of the increase in detections at this point are illustrated in Figure 3.6. The typical post-hatching period of 15 days for *F. hypoleuca*, corresponds with a clear drop in visitation detection rates (Figure 3.6), similarly after the typical post-hatching period of 20 days for *C. caeruleus* and *P. major*. This shift is reflected in the consistent fledging onset changepoint estimates and narrow credible intervals (Figure 3.5). The only nest box that did not follow the characteristic spike post-hatching was NB 39, a *P. major* nest adjacent to the River Rye. The relatively low detections were likely a result of the higher threshold for the classifier detection to reduce conflation with background noise as discussed in Chapter 2, 2.2.5 Vocalisation and Visitation Clustering and Classifier Construction, Classifier Construction.

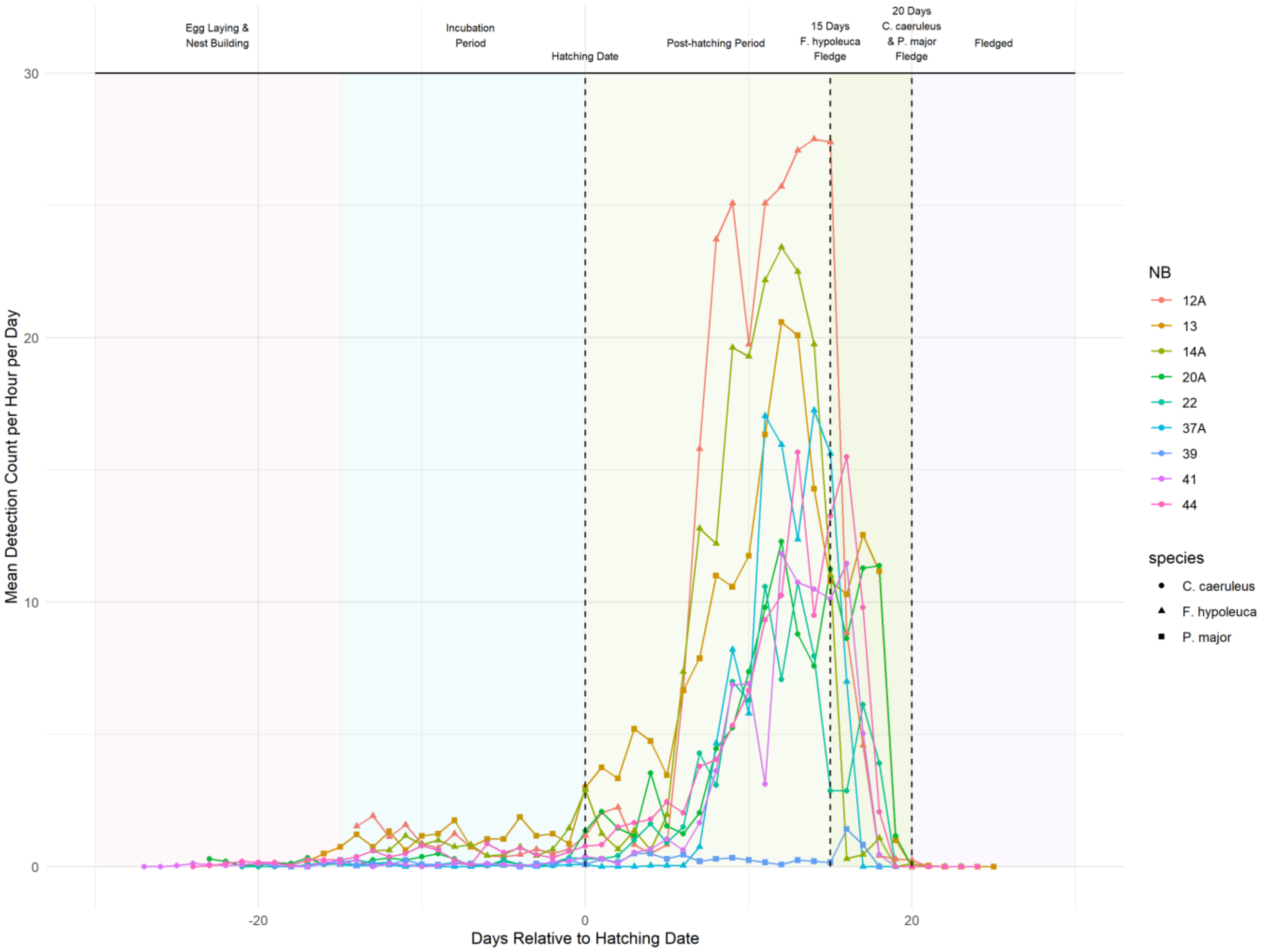

Figure 3.6, Mean hourly detection counts per day relative to hatching date for each nest box. Lines represent individual nest boxes, coloured by Nest Box ID (NB), with shapes denoting species (C. caeruleus, F. hypoleuca, P. major). Shaded regions indicate breeding stages: egg laying and nest building (pink), incubation (blue), post-hatching period (yellow), and fledging (green). Vertical dashed lines mark key breeding milestones: hatching date (day 0), 15 days post-hatching (F. hypoleuca fledging), and 20 days post-hatching (C. caeruleus and P. major fledging).

3.3.2. Vocally-derived Changepoints

The posterior model fit for each vocally-derived changepoint is provided in Appendix B Section 3.6.2. Each plot shows vocalisation patterns alongside posterior mean fits and the posterior distribution of estimated changepoints for each nest box and breeding stage transition.

Vocally-derived changepoints were estimated using vocalisation detections from recorders at *F. hypoleuca* nest boxes (Chapter 2, Table 2.1). The alignment of changepoints detected from vocalisation data compared to in-field estimates varied by breeding stage (Figure 3.7). The median deviation from the field-estimated changepoint was +9.98 days (IQR: 8.73 - 11.76) for arrival, -4.5 days (IQR: -7.5 - -2.6) for incubation, -11.5 days (IQR: -12.5 to -9.2) for hatching, and +0.4 days (IQR: -4.4 - 0.4) for fledging. Across nest boxes, credible intervals at 95% were narrowest for arrival (median width 0.92 days, IQR: 0.72 - 0.94), intermediate for hatching and fledging (0.94 days, IQR: 0.93 - 12.4 and 0.55 days, IQR: 0.3 - 6.2, respectively), and widest for incubation (4.96 days, IQR: 2.95 - 7.40).

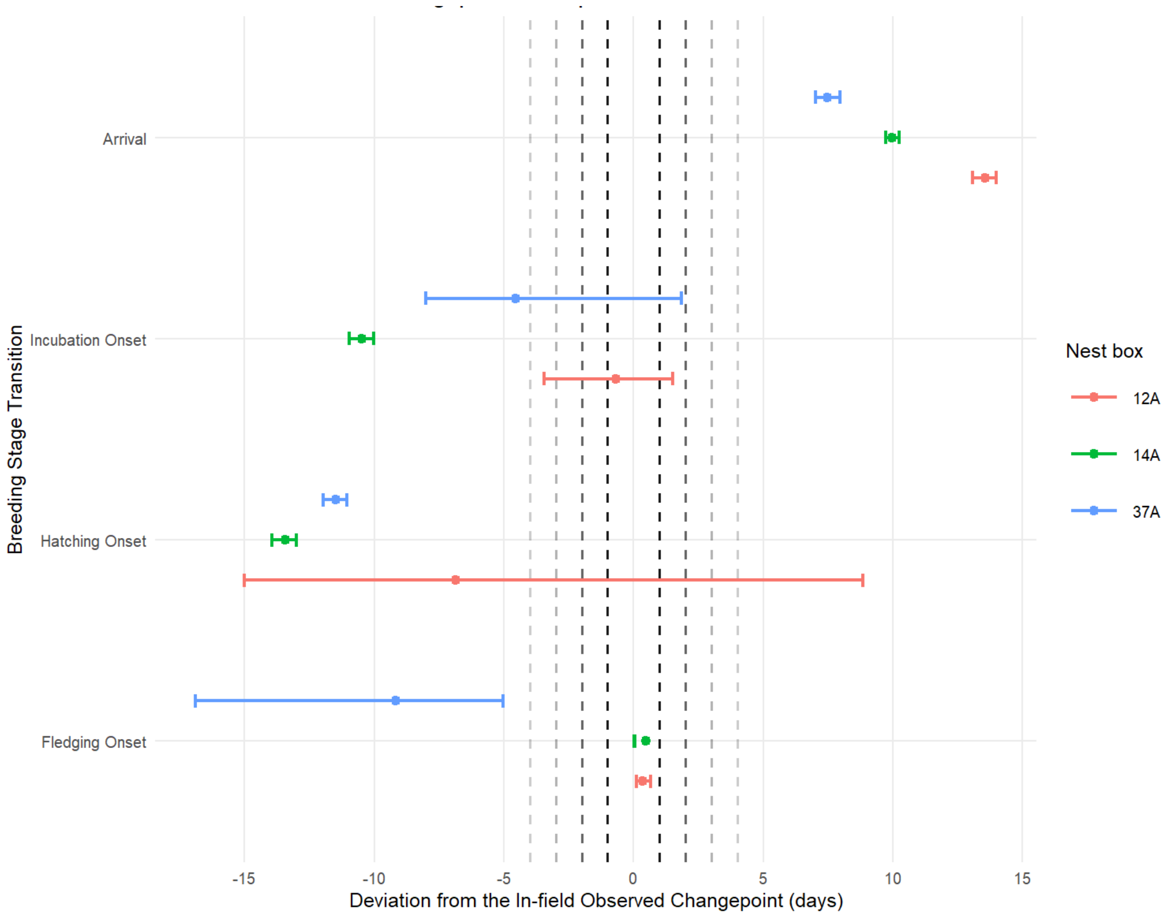

Figure 3.7 Changepoints derived from models fitted to vocalisation detections from acoustic data across all F. hypoleuca nest boxes, compared with field-estimated breeding stage transitions. The in-field arrival date estimate is taken from the first detected F. hypoleuca vocalisation from the audio data collected at each nest box. Horizontal lines showing 95% credible intervals from the changepoint model (Appendix B, Section 3.6.2). Vertical dashed lines indicate the deviations from in-field observed changepoint (0) +/- 1–5 days around the observed date (grey). Colours correspond to individual nest boxes see Chapter 2, Figure 2.2 & 2.3. Breeding stage transitions are shown separately for arrival, incubation onset, hatching onset, and fledging onset.

To visualise the distribution of vocalisation detections along the breeding timeline, the mean detection count per hour for each nest box is plotted against Julian day relative to the in-field estimate for arrival date (first vocalisation detected by the recorder) (Figure 3.8). Across all three nest boxes, the credible intervals for each estimate were narrower for arrival (Figure 3.7), indicating a distinct shift in detection rate. Accordingly, the changepoint estimate for arrival at each nest box is plotted alongside detection frequency (Figure 3.8). In each case, the changepoint analysis identified an arrival date that corresponds either to a sustained increase in vocalisations (NB37A and NB12A) or to a decline following an initial wave in detections (NB14A).

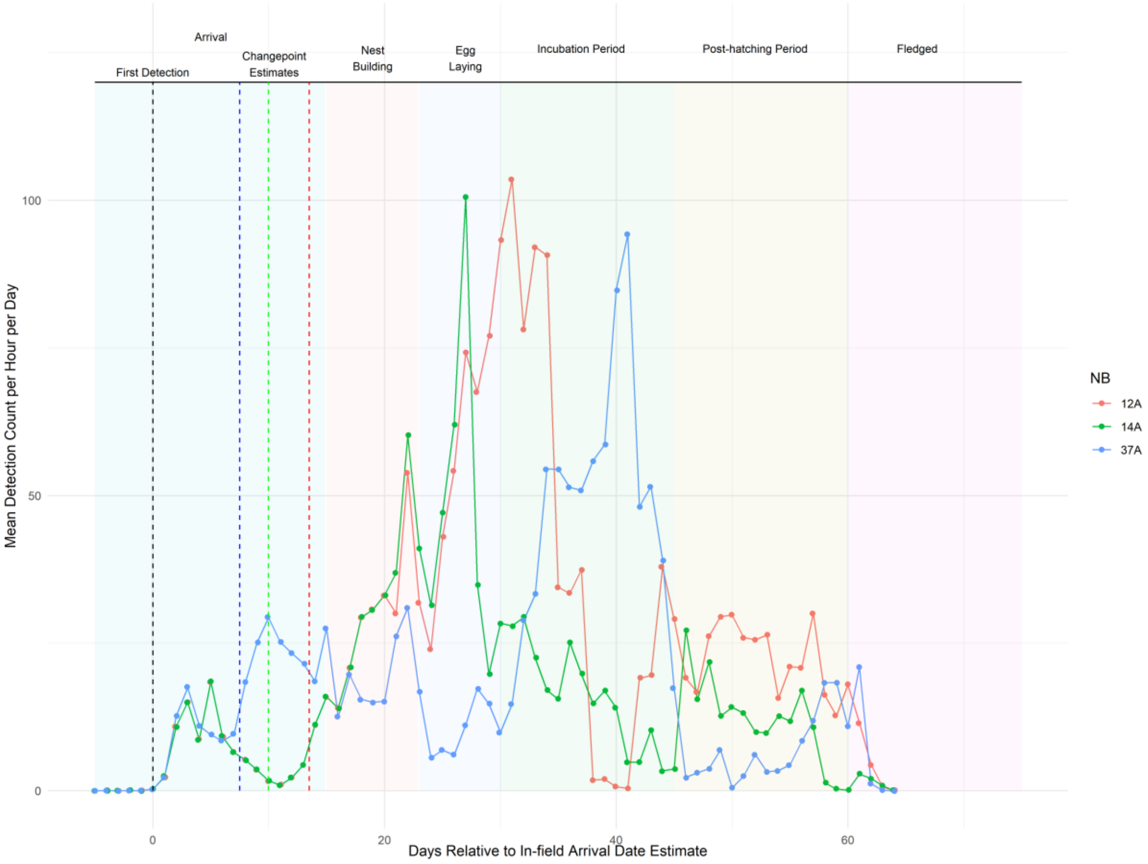

Figure 3.8 Mean hourly detection counts per day relative to in-field arrival date estimates for three nest boxes (12A, 14A, 37A). Coloured lines represent nest box IDs. The first black dashed line marks the initial vocalisation detection, while coloured dashed lines mark changepoint estimates of arrival for each nest box. Shaded regions denote key breeding stages: arrival (blue), nest building (pink), egg laying (red), incubation (blue), post-hatching period (yellow), and fledging (purple).

3.3.3. Recording Schedule Comparison - Visitation

The mixed-effects model found no evidence that recording schedule influenced the visitation-derived changepoints, as shown in Figure 3.9. Relative to constant recording, the 3-hour (Estimate = -0.66, SE = 1.35, p > 0.5), 6-hour sunrise (Estimate = -0.12, SE = 1.35, p > 0.5), and 6-hour sunset (Estimate = -0.82, SE = .35, p > 0.5) schedules all gave estimates that were within a day of the 24-hour baseline. None of these differences were statistically supported (*p* > 0.5). Overall, visitation-derived changepoints appeared robust to differences in recording schedule strategy. Full model output summary is provided in Table 3.2.

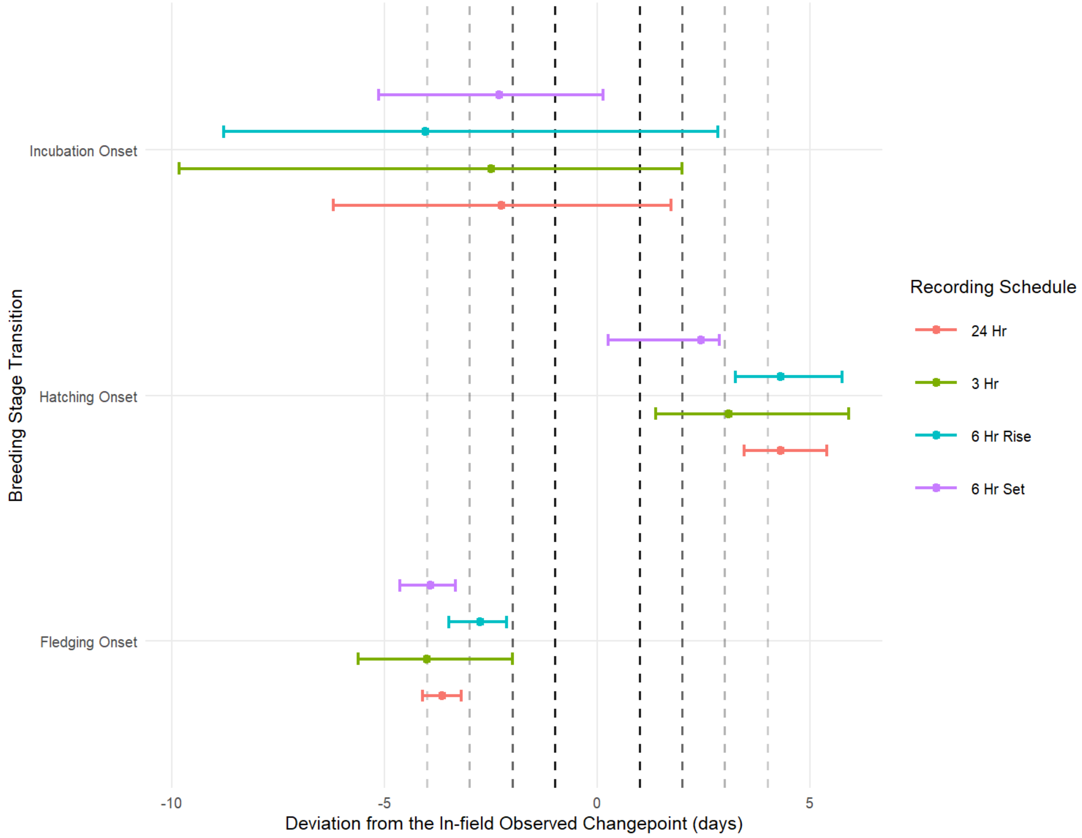

Figure 3.9, Comparison of visitation-derived changepoints across recording schedules. Points represent the mean estimated changepoints across all nest boxes for which visitations were detected (Chapter 2, Table 2.1) under four recording schedules (24 Hr - constant recording, 3 Hr - three hours post-sunrise and prior to sunset, 6 Hr Rise - six hours post-sunrise, and 6 Hr Set - six hours prior to sunset), with horizontal lines showing the mean 95% credible intervals across nest boxes. Vertical dashed lines indicate deviations from the in-field observed changepoint (0) +/- 1-5 days around the observed date (grey). Breeding stage transitions are shown for onset of incubation, hatching, and fledging.

Table 3.2 Fixed effects from the linear mixed-effects model testing the effect of recording schedule on the deviation between visitation-derived estimated and observed changepoints. Estimates are presented alongside standard errors (SE), Degrees of Freedom (df), t-values, and p-values. Non-integer degrees of freedom reflect Satterthwaite's approximation; the intercept has low degrees of freedom because it is estimated from only 9 nest boxes and 3 transitions, while predictor degrees of freedom are based on all 96 observations.

Fixed Effect	Estimate	SE	df	t-value	p-value
Intercept - 24 Hours	-0.21	2.50	3.60	-0.08	0.94
3 Hours	-0.66	1.35	82.16	-0.49	0.63
6 Hours Sunrise	-0.12	1.35	82.16	-0.09	0.93
6 Hours Sunset	-0.82	1.35	82.16	-0.61	0.55

3.3.4. Recording Schedule Comparison - Vocalisation

The mixed-effects model indicated no clear effect of recording strategy on vocalisation-derived changepoints as illustrated in Figure 3.10. Relative to constant recording, the 3-hour (Estimate = -3.77, SE = 2.04, $p > 0.05$), and 6-hour sunset (Estimate = -2.46, SE = 1.04, $p > 0.1$) schedules tended to produce earlier estimates of changepoints. While the 6-hour sunrise schedule gave near-identical results (Estimate = 0.39, SE = 2.04, $p > 0.5$). However, no differences in recording schedules were strongly supported statistically ($p > 0.05$). Full model output summary is provided in Table 3.3.

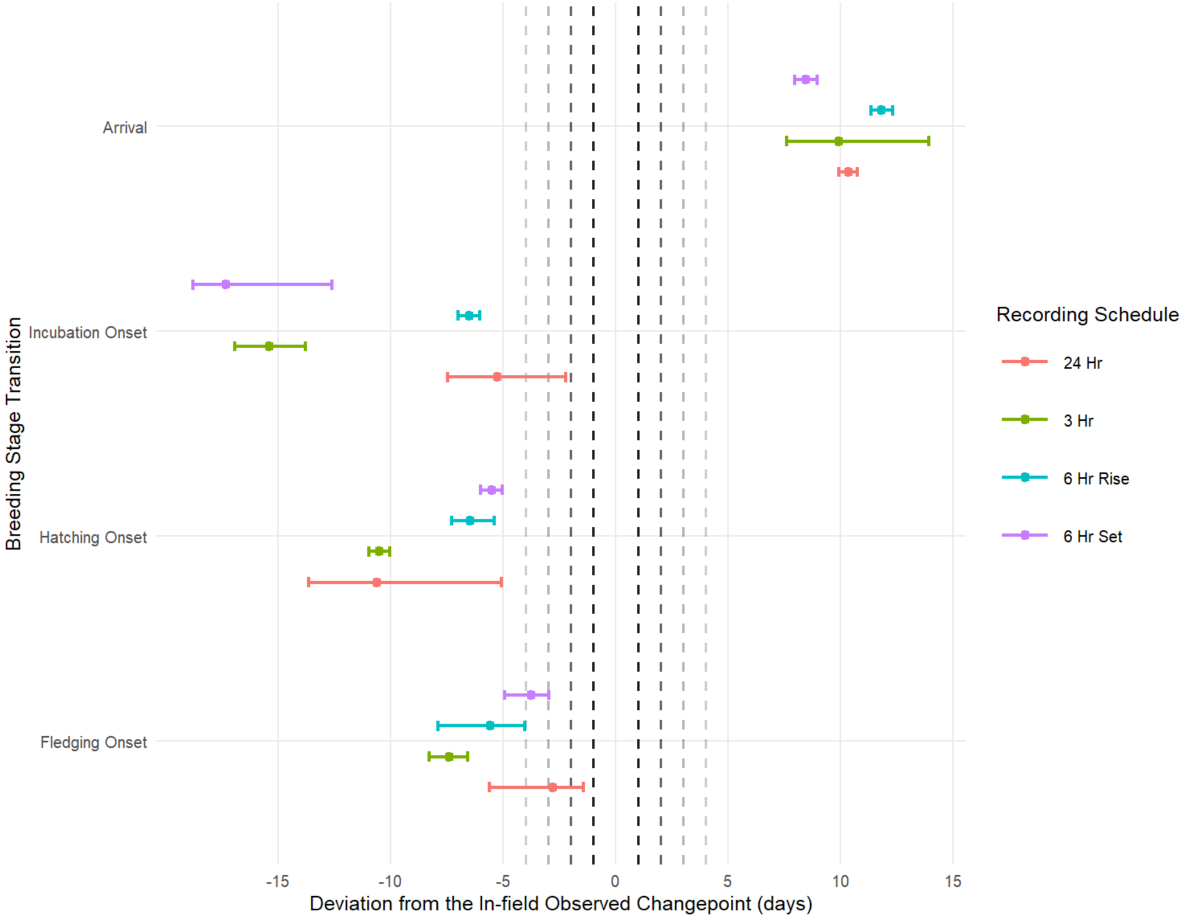

Figure 3.10, Comparison of vocalisation-derived changepoints across recording schedules. Points represent the mean estimated changepoints across all nest boxes for which vocalisations were detected (Chapter 2, Table 2.1) under four recording schedules (24 Hr - constant recording, 3 Hr - three hours post-sunrise and prior to sunset, 6 Hr Rise - six hours post-sunrise, and 6 Hr Set - six hours prior to sunset), with horizontal lines showing the mean 95% credible intervals across nest boxes. Vertical dashed lines indicate deviations from the in-field observed changepoint (0) +/- 1–5 days around the observed date (grey). Breeding stage transitions are shown for onset of arrival, incubation, hatching, and fledging.

Table 3.2, Fixed effects from the linear mixed-effects model testing the effect of recording schedule on the deviation between vocally-derived estimated and observed changepoints. Estimates are presented alongside standard errors (SE), Degrees of Freedom (df), t-values, and p-values. Non-integer degrees of freedom reflect Satterthwaite's approximation; the intercept has low degrees of freedom because it is estimated from only 3 nest boxes and 3 transitions, while predictor degrees of freedom are based on all 48 observations.

Fixed Effect	Estimate	SE	df	t-value	p-value
Intercept - Constant	-2.06	5.19	4.22	-0.397	0.71
3 Hours	-3.77	2.04	39	-1.85	0.07
6 Hours Sunrise	0.39	2.04	39	0.19	0.85
6 Hours Sunset	-2.46	1.04	39	-1.21	0.23

3.3.5. Relating Visitation Rates to Brood Size

Exploratory analysis of visitation rates in relation to breeding success was limited by the small sample size (nine nests), precluding statistical testing. Consequently, the boxplots presented rely on data from only one or two nests per brood size, resulting in high variability. Nevertheless, within-species patterns revealed varying dynamics in daily visit rates per chick across brood sizes (Figure 3.11). In *C. caeruleus* visit rates were low in small broods (six chicks; median = 1 visit per chick per day, IQR <1) but increased with brood size, reaching medians of 5.6 visits per chick per day in broods of eight and to 8.9 visits per chick for broods of nine, albeit with greater variability (IQR 16.5 and 21.1, respectively). Sample size for *C. caeruleus* declined between the hatched and fledged stages due to a failed brood, which prevented estimation of per-chick fledging rates in that case. These patterns were maintained across hatching and fledging suggesting relatively stable effort. *P. major*, showed a peak at intermediate brood sizes, with broods of 10 chicks receiving the highest rates (median = 12, IQR 25.4), while broods of 13 had lower detected visitations (median = 1.9, IQR

12.1) for both total hatched and fledged. *F. hypoleuca,* were the most variable, with the highest visitation rates at broods of six to seven chicks (median = 6.7 – 15.7), with very high interquartile ranges (56 – 81), whereas broods of eight chicks consistently received very low per-chick visit rates (median – 1.3, IQR = 28).

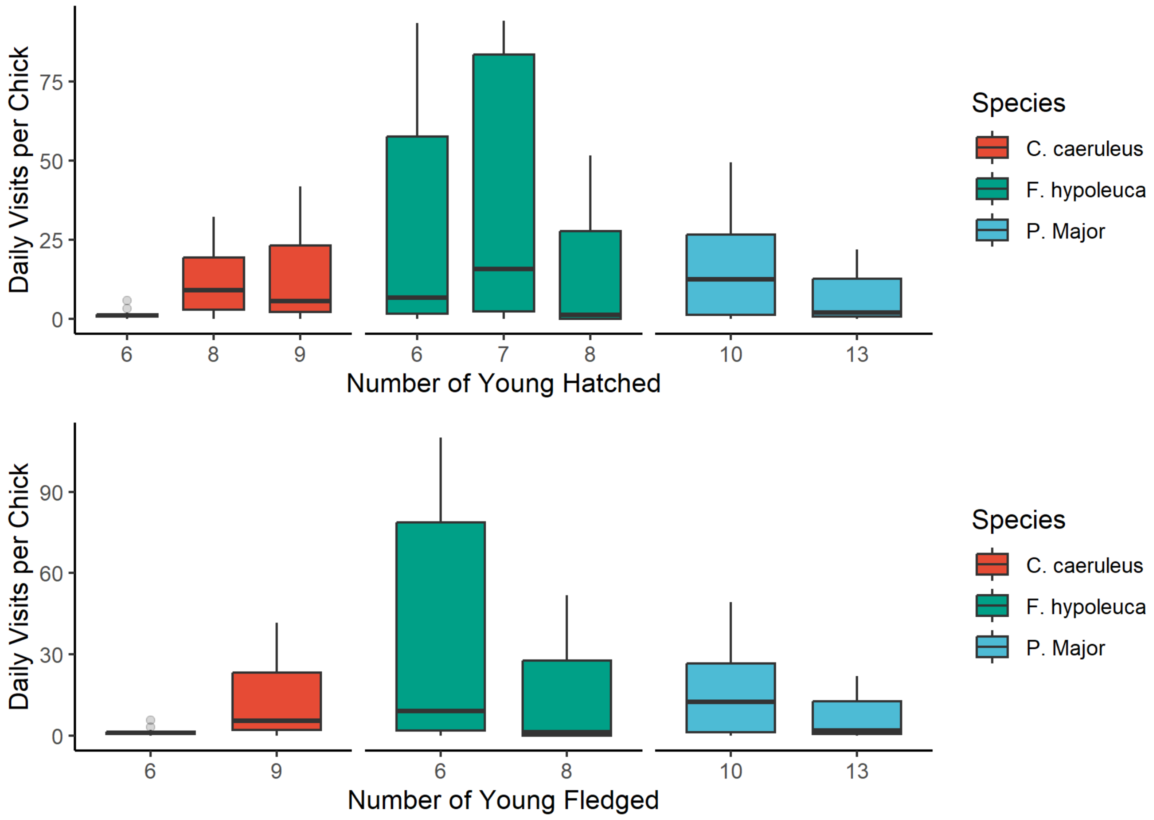

Figure 3.11 Daily parental visits per chick relative to brood size for three bird species. Boxplots show the distribution of daily visits per chick for each species across different brood sizes. Top panel: Daily visits per chick in relation to the number of young hatched. Bottom panel: Daily visits per chick in relation to the number of young fledged. For each brood size in each species, the boxplots represent a single nest, except for C. caeruleus brood size = 9, which represents two nests. Species are colour coded as follows: C. caeruleus (red), F. hypoleuca (green), P. major (blue). Boxes represent the interquartile range, horizontal lines indicate the median, whiskers show the range excluding outliers, points represent outliers.

3.4. Discussion

3.4.1. Estimating Arrival at the Nest Box

Estimated arrival dates for *F. Hypoleuca* were on average 9.9 days later than the first vocalisations detected by the audio recorder at the nest box. This delay reflects the changepoint approach, which infers arrival from a sustained increase in vocal activity rather than from isolated early detections. The early detections may represent transient individuals passing through the recording area or birds nesting at the edge of the detection radius, whereas the changepoint estimate more reliably reflects the timing of settlement by the target species. For each nest box, the changepoint estimates were before the onset of nest building (Figure 3.8), which is commonly used to approximate female arrival (Moreno et al., 2008), suggesting the method can detect a point of consistent presence at the nest box between male and female arrival.

In this study, migrant arrival dates were inferred from vocalisations detected by the audio recorders rather than visitations to the nest box, as recorders were not placed on the same trees that were eventually occupied *F. hypoleuca* boxes. For all three focal nests, however, recorders were on neighbouring boxes within 20-40 meters (Chapter 2, Figure 2.3). Given the potential detection radius of 25m for the *F. hypoleuca* vocalisation frequency range (Chapter 2, Figure 2.8) the recorders used were either within range of an occupied box, or likely capable of detecting an individual vocalising closer to the recorder elsewhere on the occupied tree or on a neighbouring tree. Enabling vocalisations detected by nearby recorders to be used as an indicator of local presence. This approach was feasible given the relatively low density of the target species in the study area, as vocalisations were highly unlikely to originate from another *F. hypoleuca* nest. Attributing detected vocalisations to the adults using the nest box will be more challenging under higher densities (see Chapter 2, Section 2.4.4.). As detecting visits to the nest box provides a more reliable measure of occupancy than vocalisations alone, when possible, future studies should test for both vocalisation and visitations. Ultimately, the estimation of arrival dates will require further scrutiny, as has been applied to existing methods such as using colour bands to monitor individual arrival dates which have been verified through geolocator tracking data (Shutt et al., 2022).

As no *F. hypoleuca* selected a recorded box for nesting, further investigation is needed to determine whether the presence of a recorder at a nest box upon arrival influences nest box choice. If found to affect nest box selection upon arrival, this would prevent the use of audio recorders to collect data on visitation events to determine arrival dates. Instead, the focus will remain on using vocalisations to infer arrival dates as demonstrated in this thesis. For now, it appears the presence of a recorder from incubation onwards does not affect productivity in *F. hypoleuca* at the study site when comparing the study year (2025) with productivity from 2018 to 2024 (Appendix A, Figure 2.A5).

3.4.2. Detecting Breeding Stage Transition

Comparisons between audio-derived changepoint estimates and in-field estimates provide insight into how well acoustic transitions can be related to breeding stage transitions. The accuracy of changepoint estimates differed between vocalisations and visitations, and across breeding stages, reflecting both the detectability of each behaviour and the extent to which behaviours change at each transition. Changes in visitation patterns consistently outperformed vocalisation patterns in aligning with the observed change in two of three comparable breeding stages (onset of hatching and fledging). Given visitation rates are a more direct representation of the changes in demands across these stages (Nowicki & Searcy, 2005; Ritchison, 2023), this outcome should be expected. As a result, future work should be focussed on detecting changes in visitation patterns at the nest box when estimating the onset of incubation, hatching, and fledging.

The accuracy of visitation-derived changepoints still varied depending upon breeding stage. For example, there are greater variation in the estimates for incubation, in comparison to hatching and fledging. Potentially a result of similar nest box entry or exit rates between pre- and post-incubation (Mariette et al., 2012). A likely explanation being egg laying does not demand consistent presence at the nest box, with visits from males and females being brief and infrequent (Slagsvold & Wiebe, 2021), whereas during incubation the female remains in the nest for extended periods, rarely leaving (Bambini et al., 2019). For hatching and fledging, changes in the rate of nest box entry and exits are more distinct due to increased demand for food provision post-hatching and the absence of young post-fledging.

The later detection of hatching relative to field observations likely reflects the onset of chick begging calls, which provide a distinct acoustic signal that is easier to identify (as discussed in Chapter 2). In this study, the changepoint estimate for the onset of hatching was detected approximately six days after the observed hatching date. If this pattern holds with larger sample sizes, it could either serve as a distinct breeding phenology marker or be corrected for hatching date by subtracting the average delay period (Romano et al., 2023). Alternatively, with a larger sample of audio recordings from nest boxes, the classifiers developed in this study could be further trained to enhance detection capabilities and potentially reduce the observed delay.

In contrast, fledging was consistently detected earlier, and with narrower credible intervals, than field observations. As fledging dates are often inferred from the typical duration of the nestling period and retrospectively confirmed during nest checks or alternatively predicted from biometric measurements taken during chick ringing (Bani Assadi et al., 2022), the earlier detections in this study may provide a more accurate and less intrusive estimate of fledging date.

The reliability of audio-derived estimates for breeding stage transitions should be evaluated relative to the uncertainty in current monitoring approaches. Nest box checks conducted in the field typically involve temporal uncertainty of approximately +/-1-2 days, as stage transitions are often inferred retrospectively based on assumed average breeding stage durations (Walker et al., 2018). In this context, the consistent alignment between audio-derived and in-field estimates of fledging suggests that acoustic monitoring can achieve a comparable level of temporal precision for certain transitions. Consequently, breeding stage timing derived from audio data appears to be at least as accurate as conventional monitoring methods that rely on standardised breeding stage length assumptions. However, this equivalence depends on transition type and improvements in classifier performance, particularly for estimating hatching date, are necessary to avoid using assumed breeding stage lengths and improve confidence across all breeding stages.

3.4.3. Exploring the Audio Effort Required to Detect Breeding Stages

As the application of using audio recorders for remote nest box monitoring had not been tested before, the study collected constant audio data, to provide a proof-of-concept for deriving changepoints from acoustic data. This dataset then permitted, through subsampling, an analysis of the impact of reduced recording schedules.

When compared with constant recording, three reduced sampling strategies at 6 hours per day showed no statistical difference in their estimated changepoints, nor the credible intervals for the estimates, across both visitation- and vocally-derived changepoints. As no difference was detected, reduced recording efforts matched the performance of constant recording in identifying arrival (based on vocalisations) as well as the onset of hatching and fledging (based on visitation patterns). This suggests that continuous monitoring is not essential and that lower-effort approaches can offer a practical yet reliable alternative for long-term studies. Such a reduced strategy would permit breeding behaviour and phenology to be audio monitored, from arrival to fledging, whilst requiring only one or two site visits to change batteries based on the total days recording in this study (77 days between 27th March and 12th June). To place these findings in the context of other commonly used audio recorders, Table 3.4 compares total recording days under a reduced recording schedule alongside the number of battery changes required. Despite their long recording capacity, the SM Mini 2 and SM4 are not cost-effective options for scalable nest box monitoring. By contrast, the SM Micro 2 and Audiomoth offer a more practical balance between cost and battery changes. Testing whether Audiomoth units can deliver data comparable to the SM Micro 2 would be valuable, given their lower cost, although they are expected to require one additional battery change. At present, the efficiency of the SM Micro 2 highlights its potential for deployment at larger scales, where logistical constraints often limit the intensity of traditional nest box monitoring (Sugai et al., 2019).

Table 3.4, Total recording hours refers to the length of continuous recording available. Days recording (6 hr/day) refers to the length of recording period without a battery change under the reduced recording schedule tested in this study. Number of battery changes was calculated by dividing the total recording days (77) by the reduced schedule, representing the number of battery changes required to maintain continuous monitoring if new batteries were installed on day 1. The Song Meter (SM) Micro 2, SM Mini 2, and SM4 (Wildlife Acoustics) used estimates based on 24 kHz stereo recordings with alkaline batteries. Tests on the Audiomoth total recording hours ranged from 16–32 kHz, also using alkaline batteries (Lapp et al., 2023). This study used the SM Micro 2 with Nickle Metal Hydride (NiMH) batteries and 64 GB or 128 GB SD cards. Comparable figures were only available for alkaline batteries. The maximum storage capacity for all equipment was 2 TB, meaning the limiting factor for total recording hours is the equipment's compatible battery type.

Equipment	Total Recording Hours	Days Recording (6 hr/day)	Number of Battery Changes	Compatible Battery Type (Alkaline)	Cost per unit (£)
SM Micro 2	280	46	1	4 x AA	150
Audiomoth	189 - 224	31 - 37	2	3 x AA	72
SM Mini 2	530	88	0	8 x AA	365
SM 4	440	73	1	4 x D-cell	585

3.4.4. Scalability of Changepoint Methods

The Bayesian changepoint regression demonstrated that acoustic detections could provide quantitative estimates of breeding stage transitions, along with their associated uncertainty. Testing each transition period against a small set candidate models (varying in number of changepoints, slopes, and autoregressive terms) meant that we could derive changepoints that closely aligned with field-observed transitions across nest boxes.

In this exploratory analysis, changepoint estimates were guided by field observations of breeding stage transitions, which constrained the level of automation. However, this requirement may lessen as methods develop. Once patterns of behavioural activity across the breeding period are better characterised, changepoint models could be adapted to detect multiple transitions directly from the data, rather than relying on in-field observations to subset stages as was done in this analysis.

This study treated vocalisations and visitations in isolation when estimating changepoints. However, integrating information from multiple behavioural signals may provide more accurate or robust estimates of breeding stage transitions. Future research could therefore explore combining changepoints across different acoustic behaviours to refine the dating of key breeding events.

3.4.5. Relating Visitation to Breeding Success

Given that the primary purpose of parental visits to the nest box after hatching is food provisioning (Ritchison, 2023), a positive linear relationship between visitation rates and brood size has been reported in passerines (Madden et al., 2022; Sanz & Tinbergen, 1999). However, this relationship is not apparent in exploratory Figure 3.11, for any of the three species across the nine boxes. For *C. caeruleus,* the very low detection rate in the brood of 6 chicks is likely an artefact of the adjusted detection threshold in a nest close to the river, where low-frequency noise interfered with visit detection (see Chapter 2, Section 2.2.5. – Classifier Construction). When this case is excluded, there appears to be no clear trend between visitations per chick and either measure of brood size. In *F. hypoleuca*, a slight negative trend is observed when comparing visitation rates with fledging number, which is not present when considering total number of hatched chicks. For *P. major*, a slight negative trend exists between visitations per chick and both measures of brood size.

In the case of *P. major,* where previous studies have found a strong relationship between visitation rates and brood size (Sanz & Tinbergen, 1999), the absence of this pattern in the audio data may result from either limitations in the detection method or the small sample size. These same factors could also explain the patterns observed in *C. caeruleus* or *F. hypoleuca*. However, because prior knowledge of visitation-brood size relationships in these species was unknown, the effect may be weaker or less consistent. The extent to which this relationship varies among species could reflect the trade-off between parental investment in offspring and the costs to their own condition. Since larger broods require more resources, a tapering of visitation rates at higher brood sizes might be expected (Bambini et al., 2019; Mariette et al., 2012). In that case, it may be reasonable to expect a lower per chick visitation rate at higher brood sizes.

When interpreting these patterns across nests, it is also important to recognise that observed visitation rates are influenced not only by parental behaviour but also by the limitations of the detection method. The influence of surrounding soundscapes on detection reliability has reduced robustness of comparisons across nests in this study. One way to resolve this limitation would be to incorporate nest-level environmental data such as background noise levels, habitat structure, and micro-climate data, as covariates in models of visitation rate (Bambini et al., 2019). These steps would allow separation of true visitation behaviour from variation in detection reliability, thereby improving the robustness of between-nest comparisons.

Similarly, this could include changes in resource availability between nests which would likely contribute to visitation frequency, with nests in poorer habitats less abundant in resources necessitating a longer foraging time for adults (Madden et al., 2022). Accounting for changes in climactic conditions using micro-climate loggers will also help explain differences in behaviour in response to local differences in temperature and rainfall between nests (Scholl & Hille, 2020; Shutt et al., 2022)

Using models that incorporate environmental covariates, require larger sample sizes than were available in this study. With only nine nests monitored, complex models risk overfitting and producing unstable estimates, meaning inferences had to rely on simpler comparisons. This constraint also limited any meaningful statistical inference on the relationship between visitation rates and chick growth, an initial objective of this study for which sample size was insufficient (see Appendix B - Figure 3.B8). For now, a continuation of recording audio at nest boxes with in-person monitoring to provide ground-truthing is required to help clarify the relationship between nest visitation and brood sizes.

3.4.6. Behavioural Signatures of Nest Activity and Success or Failure

Audio data may be able to determine binary metrics for breeding success; whether a nest is active and if a brood has succeeded or failed based on vocalisation and visitation frequencies. For example, in Figure 3.8, mean hourly vocalisation detections peaked during egg-laying and early incubation and dropped during incubation and prior to hatching across all nest boxes. This pattern could reflect behavioural shifts in response to a change in breeding stage and provide a characteristic pattern indicative of nest activity. During egg-laying and incubation, parents may communicate more frequently, in *P. major* for example, a male will call to the incubating female to indicate when it is safe to leave the nest for food (Viigipuu et al., 2023). Close to hatching, however, adults may reduce vocal activity to avoid drawing attention to vulnerable eggs and chicks (Yoon et al., 2016), while their time and energy are redirected toward provisioning (Ritchison, 2023). As the pattern was observed across all three nest boxes, this option seems promising, however additional nests will need to be monitored to fully evaluate this approach for determining nest activity.

Visitation patterns plotted along the breeding timeline could also indicate brood success or failure. If visitations persist for the full duration of the typical post-hatching period (well established for cavity-nesting species (Walker et al., 2018)), it is likely that parents continued visiting the nest until fledging. In contrast, if visitation detections declined prior to this expected timeline, it suggests that parents were no longer regularly attending the nest, potentially due to brood failure. This pattern was evident in the present study, where the *C. caeruleus* nest in box 22 failed and visitation rates declined earlier in the breeding timeline compared with other *C. caeruleus* nests (Figure 3.12).

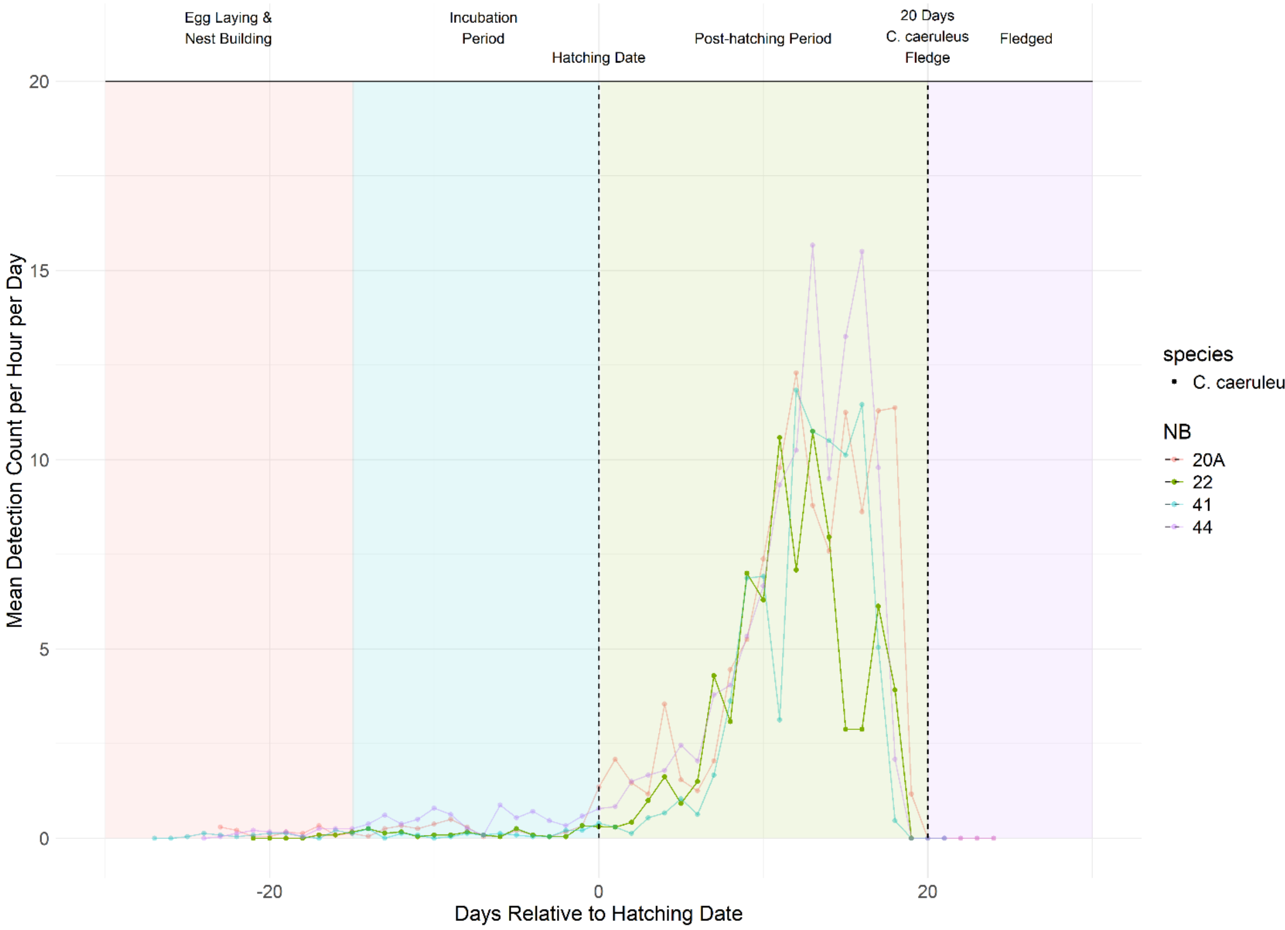

Figure 3.12 Mean hourly detection counts per day relative to hatching date for C. caeruleus nest boxes. Lines represent individual nest boxes, coloured by nest box ID (NB), with shapes denoting species (C. caeruleus). The nest box with the failed brood, 22, is in bold green to show differences in the visitation rates over time when a brood fails compared to successful C. caeruleus boxes (transparent). Shaded regions indicate breeding stages: egg laying and nest building (pink), incubation (blue), post-hatching period (yellow), and fledging (green). Vertical dashed lines mark key breeding milestones: hatching date (day 0), and 20 days post-hatching (C. caeruleus fledging).

3.4.7. Wider Applications

Based on the evidence provided in this study, audio recorders at the nest box can estimate arrival dates and breeding stage transitions with minimal disturbance to the nest and with much reduced researcher field effort.

This approach, either used in combination with a reduced form of traditional nest box monitoring or independent of it, has the potential to increase the scale of monitoring of breeding phenology estimates in cavity-nesting species. Beyond methodological efficiency, such tools could support assessments of shifting breeding phenology under climate change (Newson et al., 2016; Nicolau et al., 2021) and, when paired with micro-climate sensors, could provide new insights into how migrants select breeding locations and adjust behaviour in response to local environmental conditions (Shutt et al., 2022).

Estimating arrival dates of migrants using remote, time- and cost-effective methods is of particular interest. Studies that track individuals to link migration with breeding phenology often rely on costly tracking devices or labour-intensive colour-ringing and repeated monitoring of individuals from the preceding year or upon arrival (Bejarano & Jahn, 2018; Bell et al., 2022). By contrast, understanding general migration phenology does not require individual tracking but is typically based on field reports of first arrivals – a method that is also labour-intensive (Newson et al., 2016; Nicolau et al., 2021).

Acoustic monitoring offers a potential solution, providing consistent and scalable detection of arrival patterns without the need for extensive field observations. Establishing this methodology at larger scales could refine and expand data on monitoring of populations, as well as adding arrival date estimation to the data that can be derived from nestbox schemes. This could help refine phenological studies. For example, incorporating such data into ongoing novel research that simulates individual or population-level migrations, could improve simulations (Howard et al., 2024).

The outcomes of this study suggest that remote nest box audio recorders can provide breeding phenology estimates, and while they are unlikely to yield reliable estimates of brood size as an indicator of breeding success (Figure 8 and 9), they may provide a binary metric of nest outcome, distinguishing successful from failed broods. For now, audio recorders at nest boxes should be used in combination with traditional nest monitoring, both to further evaluate audio-derived estimates of breeding success and to further our understanding of the relationship between arrival dates and breeding success.

3.5. References

Bambini, G., Schlicht, E., & Kempenaers, B. (2019). Patterns of female nest attendance and male feeding throughout the incubation period in Blue Tits Cyanistes caeruleus. *Ibis*, *161*(1), 50–65. https://doi.org/10.1111/IBI.12614

Bani Assadi, S., McKinnon, E. A., Cheskey, E. D., & Fraser, K. C. (2022). Does hatch date set the clock? Timing of post-fledging movements for families of a colonially breeding, long-distance migratory songbird. *Journal of Avian Biology*, *2022*(4). https://doi.org/10.1111/JAV.02766

Beckage, B., Joseph, L., Belisle, P., Wolfson, D. B., & Platt, W. J. (2007). Bayesian change-point analyses in ecology. *New Phytologist*, *174*(2), 456–467. https://doi.org/10.1111/J.1469-8137.2007.01991.X

Bejarano, V., & Jahn, A. E. (2018). Relationship between arrival timing and breeding success of intra-tropical migratory Fork-tailed Flycatchers (Tyrannus savana). *Journal of Field Ornithology*, *89*(2), 109–116. https://doi.org/10.1111/jofo.12251

Bell, F., Bearhop, S., Briedis, M., El Harouchi, M., Bell, S. C., Castello, J., & Burgess, M. (2022). Geolocators reveal variation and sex-specific differences in the migratory strategies of a long-distance migrant. *Ibis*, *164*(2), 451–467. https://doi.org/10.1111/ibi.13017

Bennett, S., Williamson, L., Hernández-González, M., Denton, T., Laber, R., Deakin, Z., Bolton, M., Manilow, E., & Wilson, L. J. (2025). Bioacoustics as a Measure of Population Size and Breeding Success of European Storm Petrels Hydrobates pelagicus. *Ecology and Evolution*, *15*(8). https://doi.org/10.1002/ECE3.71893

Berger-Tal, O., Polak, T., Oron, A., Lubin, Y., Kotler, B. P., & Saltz, D. (2011). Integrating animal behavior and conservation biology: A conceptual framework. *Behavioral Ecology*, *22*(2), 236–239. https://doi.org/10.1093/beheco/arq224

British Trust for Ornithology. (n.d.). *NRS Code of Conduct | BTO*. Retrieved September 13, 2025, from https://www.bto.org/get-involved/volunteer/projects/nrs/taking-part/coc

Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2014). Temporal and environmental influences on the vocal behaviour of a nocturnal bird. *Journal of Avian Biology*, *45*(6), 591–599. https://doi.org/10.1111/JAV.00411

Ducrettet, M., Forget, P. M., Ulloa, J. S., Yguel, B., Gaucher, P., Princé, K., Haupert, S., & Sueur, J. (2020). Monitoring canopy bird activity in disturbed landscapes with automatic recorders: A case study in the tropics. *Biological Conservation*, *245*, 108574. https://doi.org/10.1016/J.BIOCON.2020.108574

Dunn, P. O., & Winkler, D. (2010). *Effects of climate change on timing of breeding and reproductive success in birds*.

Gabry, J. (2020). *Package "loo" Type Package Title Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models*. https://doi.org/10.1007/s11222-016-9696-4

García, J., Laiolo, P., & Suárez-Seoane, S. (2024). Song complexity is associated with habitat quality in an upland passerine. *Ibis*, *166*(2), 682–693. https://doi.org/10.1111/IBI.13257

Hereward, H. F. R., Facey, R. J., Sargent, A. J., Roda, S., Couldwell, M. L., Renshaw, E. L., Shaw, K. H., Devlin, J. J., Long, S. E., Porter, B. J., Henderson, J. M., Emmett, C. L., Astbury, L., Maggs, L., Rands, S. A., & Thomas, R. J. (2021). Raspberry Pi nest cameras: An affordable tool for remote behavioral and conservation monitoring of bird nests. *Ecology and Evolution*, *11*(21), 14585–14597. https://doi.org/10.1002/ece3.8127

Howard, C., Mason, T. H. E., Baillie, S. R., Border, J., Hewson, C. M., Houston, A. I., Pearce-Higgins, J. W., Bauer, S., Willis, S. G., & Stephens, P. A. (2024). Explaining and predicting animal migration under global change. *Diversity and Distributions*, *30*(2). https://doi.org/10.1111/ddi.13797

Kellermann, J. L., & van Riper, C. (2015). Detecting mismatches of bird migration stopover and tree phenology in response to changing climate. *Oecologia*, *178*(4), 1227–1238. https://doi.org/10.1007/s00442-015-3293-7

Kern, M., Bacon, W., Long, D., & Cowie, R. J. (2001). Possible Roles for Corticosterone and Critical Size in the Fledging of Nestling Pied Flycatchers. *Https://Doi.Org/10.1086/322927*, *74*(5), 651–659. https://doi.org/10.1086/322927

Kristensen, N. P., Johansson, J., Jonzén, N., & Smith, H. G. (2018). Evolution of resident bird breeding phenology in a landscape with heterogeneous resource phenology and carryover effects. *Evolutionary Ecology*, *32*(5), 509–528. https://doi.org/10.1007/S10682-018-9951-6/FIGURES/4

Kuznetsova, A., Brockhoff, P. B., Christensen, R., & Jensen, S. (2025). *Package "lmerTest" Type Package Title Tests in Linear Mixed Effects Models*. https://doi.org/10.18637/jss.v082.i13

Lapp, S., Stahlman, N., & Kitzes, J. (2023). A Quantitative Evaluation of the Performance of the Low-Cost AudioMoth Acoustic Recording Unit. *Sensors*, *23*(11). https://doi.org/10.3390/s23115254

Lindeløv, J. K. (2020). *mcp: An R Package for Regression With Multiple Change Points*. https://doi.org/10.31219/OSF.IO/FZQXV

Madden, S. A., McDermott, M. T., & Safran, R. J. (2022). Visitation rate, but not foraging range, responds to brood size manipulation in an aerial insectivore. *Behavioral Ecology and Sociobiology*, *76*(10), 1–16. https://doi.org/10.1007/S00265-022-03244-Z/TABLES/3

Mariette, M. M., Griffi, S. C., Mariette, M. M., & Griffi, S. C. (2012). Nest visit synchrony is high and correlates with reproductive success in the wild Zebra finch Taeniopygia guttata. *Journal of Avian Biology*, *43*(2), 131–140. https://doi.org/10.1111/J.1600-048X.2012.05555.X

Moreno, J., Martínez, J., Corral, C., Lobato, E., Merino, S., Morales, J., Martínez-de la Puente, J., & Tomás, G. (2008). Nest construction rate and stress in female Pied Flycatchers Ficedula hypoleuca. *Acta Ornithologica*, *43*(1), 57–64. https://doi.org/10.3161/000164508X345338

Moreno, J., Potti, J., & Merino, S. (1997). Parental Energy Expenditure and Offspring Size in the Pied Flycatcher Ficedula hypoleuca. *Oikos*, *79*(3), 559. https://doi.org/10.2307/3546900

Newson, S. E., Moran, N. J., Musgrove, A. J., Pearce-Higgins, J. W., Gillings, S., Atkinson, P. W., Miller, R., Grantham, M. J., & Baillie, S. R. (2016). Long-term changes in the migration phenology of UK breeding birds detected by large-scale citizen science recording schemes. *Ibis*, *158*(3), 481–495. https://doi.org/10.1111/IBI.12367

Nicolau, P. G., Burgess, M. D., Marques, T. A., Baillie, S. R., Moran, N. J., Leech, D. I., & Johnston, A. (2021). Latitudinal variation in arrival and breeding phenology of the pied flycatcher Ficedula hypoleuca using large-scale citizen science data. *Journal of Avian Biology*, *52*(2). https://doi.org/10.1111/jav.02646

Nicolaus, M., Bouwman, K. M., & Dingemanse, N. J. (2008). Effect of PIT Tags on the Survival and Recruitment of Great Tits Parus major.

Https://Doi.Org/10.5253/078.096.0215, *96*(2), 286–292. https://doi.org/10.5253/078.096.0215

Nowicki, S., & Searcy, W. A. (2005). Song and Mate Choice in Birds: How The Development of Behavior Helps us Understand Function. *The Auk*, *122*(1), 1–14. https://doi.org/10.1093/AUK/122.1.1

Oestreich, W. K., Oliver, R. Y., Chapman, M. S., Go, M. C., & McKenna, M. F. (2024). Listening to animal behavior to understand changing ecosystems. *Trends in Ecology and Evolution*, *39*(10), 961–973. https://doi.org/10.1016/J.TREE.2024.06.007/ASSET/F890C929-11B9-4C29-84DB-9153F5066AED/MAIN.ASSETS/GR3.JPG

Patel, S. H., Morreale, S. J., Panagopoulou, A., Bailey, H., Robinson, N. J., Paladino, F. V, Margaritoulis, D., Spotila, J. R., Patel, C. :, Morreale, S. J., Panagopoulou, A., Bailey, H., Robinson, N. J., Paladino, F. V, Margaritoulis, D., & Peters, D. P. C. (2015). Changepoint analysis: a new approach for revealing animal movements and behaviors from satellite telemetry data. *Ecosphere*, *6*(12), 1–13. https://doi.org/10.1890/ES15-00358.1

Penteriani, V., Delgado, M. D. M., Stigliano, R., Campioni, L., & Sánchez, M. (2014). Owl dusk chorus is related to the quality of individuals and nest-sites. *Ibis*, *156*(4), 892–895. https://doi.org/10.1111/IBI.12178

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna,

Ritchison, G. (2023). *Avian Reproduction: Post-hatching Parental Care and Brood Parasitism*. 2383–2506. https://doi.org/10.1007/978-3-031-14852-1_19

Romano, A., Garamszegi, L. Z., Rubolini, D., & Ambrosini, R. (2023). Temporal shifts in avian phenology across the circannual cycle in a rapidly changing climate: A global meta-analysis. *Ecological Monographs*, *93*(1), e1552. https://doi.org/10.1002/ECM.1552;WEBSITE:WEBSITE:ESAJOURNALS;JOURNAL:JOURNAL:15577015;WGROUP:STRING:PUBLICATION

Sanz, J. J., & Tinbergen, J. M. (1999). Energy expenditure, nestling age, and brood size: an experimental study of parental behavior in the great tit Parus major. *Behavioral Ecology*, *10*(5), 598–606. https://doi.org/10.1093/BEHECO/10.5.598

Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., Woodruff, J., Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., & Woodruff, J.

(2020). *Journal of Fish and Wildlife Management | www.fwspubs.org*. *11*. https://doi.org/10.3996/102019-JFWM-083

Schöll, E. M., & Hille, S. M. (2020). Heavy and persistent rainfall leads to brood reduction and nest failure in a passerine bird. *Journal of Avian Biology*, *51*(7). https://doi.org/10.1111/JAV.0241

Shutt, J. D., Bell, S. C., Bell, F., Castello, J., El Harouchi, M., & Burgess, M. D. (2022). Territory-level temperature influences breeding phenology and reproductive output in three forest passerine birds. *Oikos*, *2022*(8). https://doi.org/10.1111/oik.09171

Slagsvold, T., & Wiebe, K. L. (2021). Egg covering in cavity nesting birds may prevent nest usurpation by other species. *Behavioral Ecology and Sociobiology*, *75*(8), 1–10. https://doi.org/10.1007/S00265-021-03045-W/FIGURES/5

Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., & Llusia, D. (2019). Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience*, *69*(1), 15–25. https://doi.org/10.1093/BIOSCI/BIY147

Surmacki, A., & Podkowa, P. (2022). The use of trail cameras to monitor species inhabiting artificial nest boxes. *Ecology and Evolution*, *12*(2), e8550. https://doi.org/10.1002/ECE3.8550

Teixeira, D., Linke, S., Hill, R., Maron, M., & van Rensburg, B. J. (2022). Fledge or fail: Nest monitoring of endangered black-cockatoos using bioacoustics and open-source call recognition. *Ecological Informatics*, *69*, 101656. https://doi.org/10.1016/J.ECOINF.2022.101656

Teixeira, D., Maron, M., & van Rensburg, B. J. (2019). Bioacoustic monitoring of animal vocal behavior for conservation. *Conservation Science and Practice*, *1*(8), e72. https://doi.org/10.1111/CSP2.72

Upham-Mills, E. J., Reimer, J. R., Haché, S., Lele, S. R., & Bayne, E. M. (2020). Can singing rate be used to predict male breeding status of forest songbirds? A comparison of three calibration models. *Ecosphere*, *11*(1), e03005. https://doi.org/10.1002/ECS2.3005

Vehtari, A., Gelman, · Andrew, Gabry, J., & Gelman, A. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*, *27*, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Viigipuu, R., Mägi, M., & Tilgar, V. (2023). Great tits alter incubation behaviour in noisy environments. *Journal of Ethology*, *41*(1), 39–46. https://doi.org/10.1007/s10164-022-00765-y

Walker, R. H., Robinson, R. A., Barimore, C. J., Blackburn, J. R., Barber, L. J., Clark, J. A., Clewley, G. D., Evans, H. E., Green, R. M., de Palacio, D. X., Grantham, M. J., Griffin, B. M., Schäfer, S., & Leech, D. I. (2018). Bird ringing and nest recording in Britain and Ireland in 2017. *Ringing and Migration*, *33*(2), 99–145. https://doi.org/10.1080/03078698.2018.1737454;JOURNAL:JOURNAL:TRAM20;PAGE:STRING:ARTICLE/CHAPTER

Wong, B. B. M., & Candolin, U. (2015). Behavioral responses to changing environments. *Behavioral Ecology*, *26*(3), 665–673. https://doi.org/10.1093/BEHECO/ARU183

Yoon, J., Kim, B. S., Joo, E. J., & Park, S. R. (2016). Nest predation risk influences a cavity-nesting passerine during the post-hatching care period. *Scientific Reports*, *6*(1), 1–7. https://doi.org/10.1038/SREP31989;SUBJMETA

3.6. Appendix – B

3.6.1. Visitation-Derived Changepoints - Posterior Model Fits

Estimating Changepoints for the Onset of Incubation

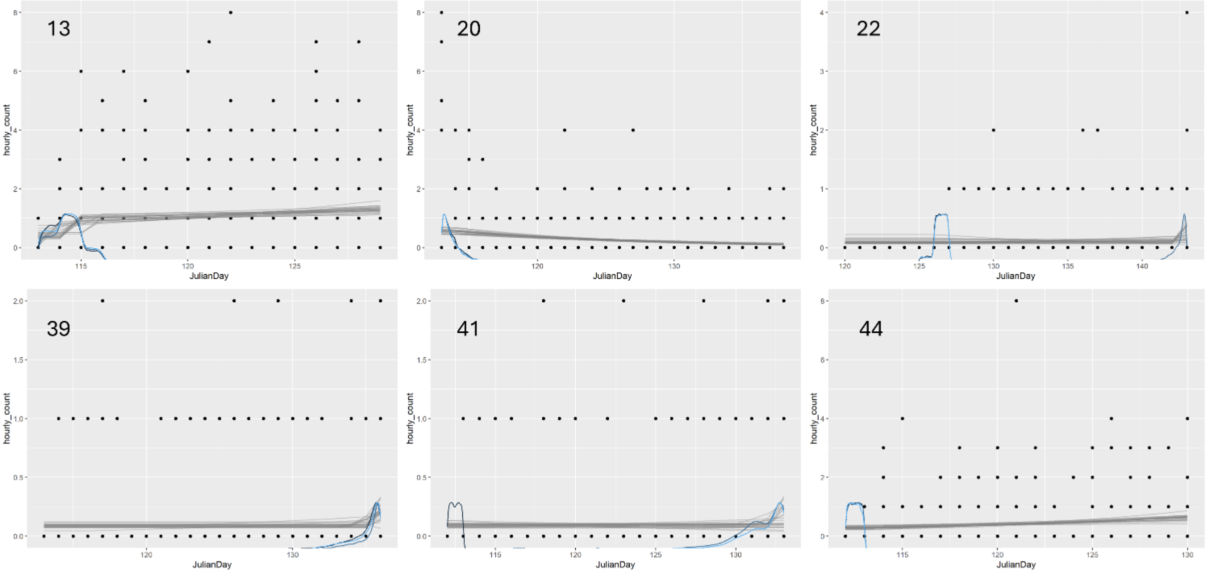

Figure 3.B1, Posterior model fits of hourly visitation detection counts across Julian days for individual nest boxes. The breeding periods included in this analysis are approx. 7 days before and 14 days after the onset of incubation. For posterior distributions with multiple estimated changepoints, the changepoint closest to the target transition (onset of incubation) was chosen Black points represent observed hourly counts, black lines show posterior mean fit, and blue lines indicate the posterior distribution of estimated changepoints. Each panel corresponds to a different nest box, ID shown in top left corner.

Estimating Changepoints between Incubation and Post-hatching

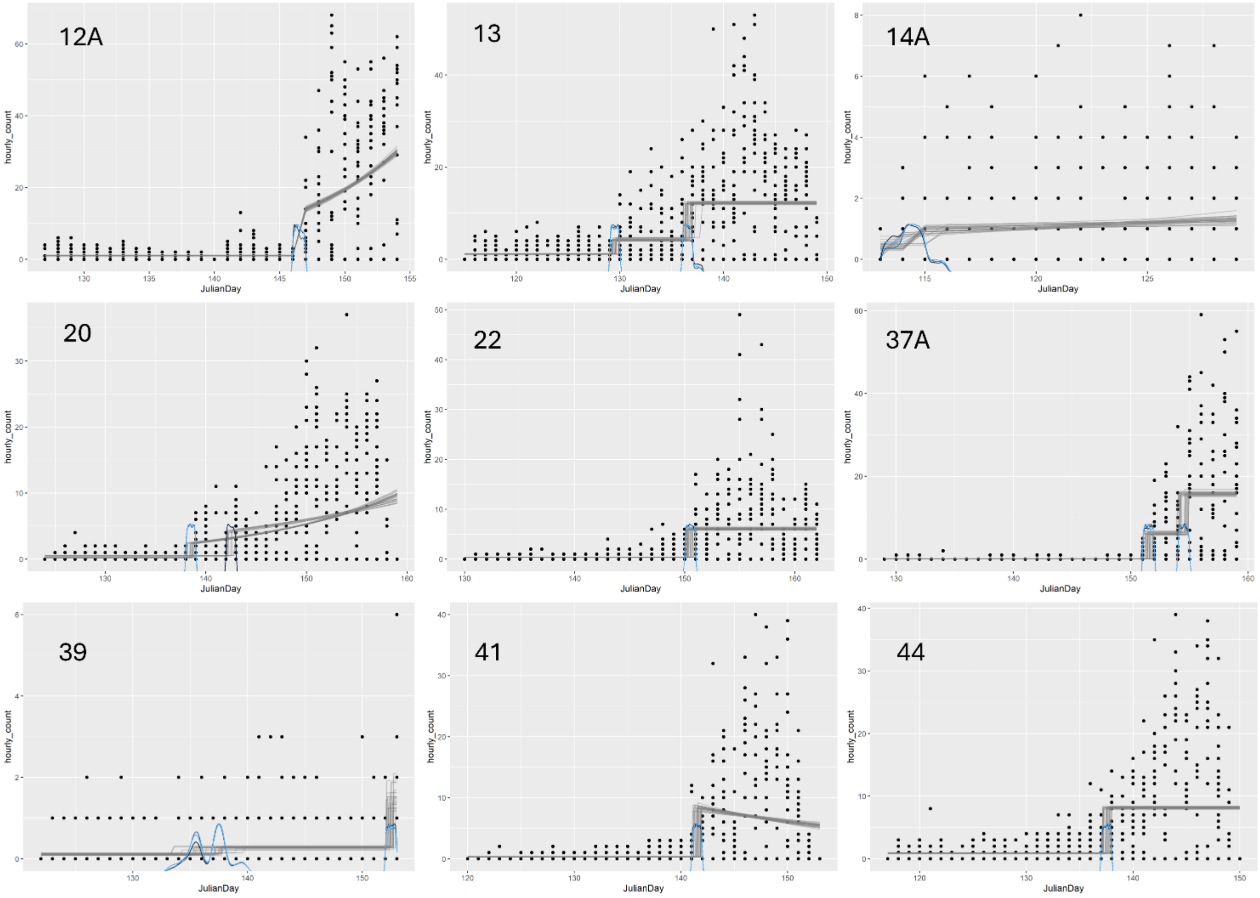

Figure 3.B2, Posterior model fits of hourly visitation detection counts across Julian days for individual nest boxes. The breeding periods included in this analysis are approx. 7 days before and 14 to 21 days after the onset of hatching, depending upon the length of the post-hatching period. For posterior distributions with multiple estimated changepoints, the changepoint closest to the target transition (onset of hatching) was chosen. Black points represent observed hourly counts, black lines show posterior mean fit, and blue lines indicate the posterior distribution of estimated changepoints. Each panel corresponds to a different nest box, ID shown in top left corner.

Estimating Changepoints between Post-hatching and Fledging

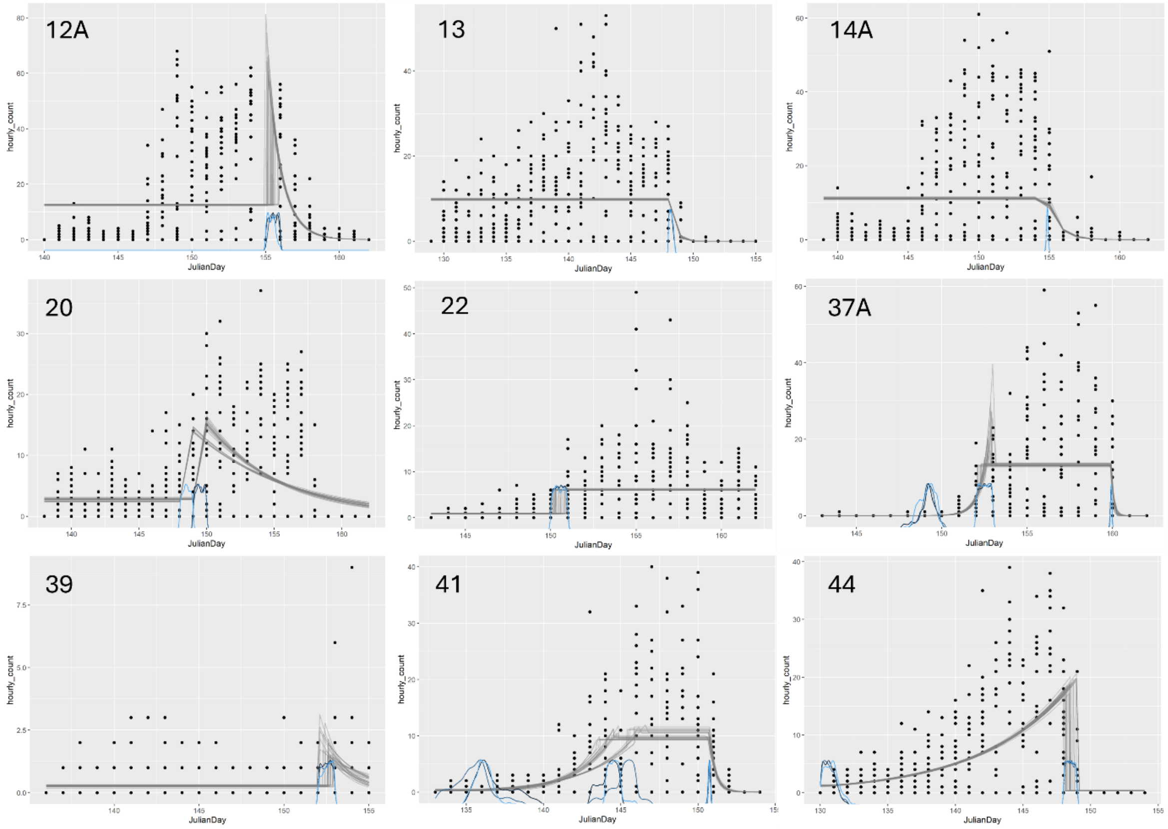

Figure 3.B3, Posterior model fits of hourly visitation detection counts across Julian days for individual nest boxes. The breeding periods included in this analysis are approx. 14 to 21 days before the onset of hatching to approx. 5 days post-fledging, depending upon the length of the species post-hatching period. For posterior distributions with multiple estimated changepoints, the changepoint closest to the target transition (onset of fledging) was chosen. Black points represent observed hourly counts, black lines show posterior mean fit, and blue lines indicate the posterior distribution of estimated changepoints. Each panel corresponds to a different nest box, ID shown in top left corner.

3.6.2. Vocally-Derived Changepoints - Posterior Model Fits

Posterior Model Fits - Estimating Changepoints During Arrival at the Nest box

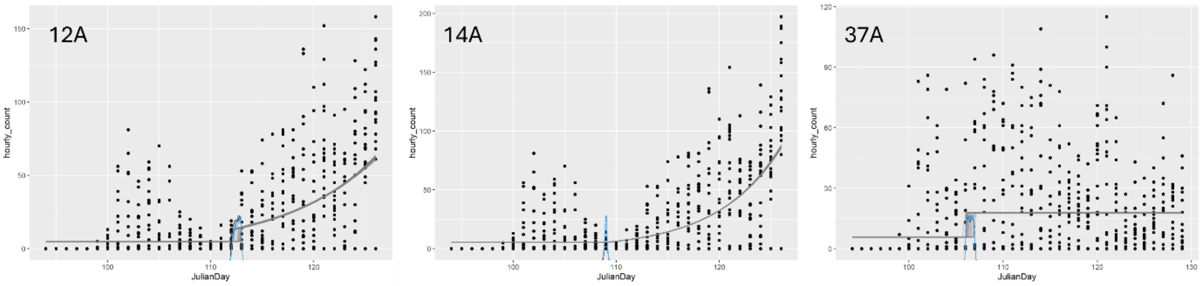

Figure 3.B4, Posterior model fits of hourly vocal detection counts across Julian days for individual nest boxes. The breeding periods included in this analysis are approx. 7 days before the first vocalisation detected to approx. 21 to 28 days after. Black points represent observed hourly counts, black lines show posterior mean fit, and blue lines indicate the posterior distribution of estimated changepoints. Each panel corresponds to a different nest box, ID shown in top left corner.

Estimating Changepoints for the Onset of Incubation

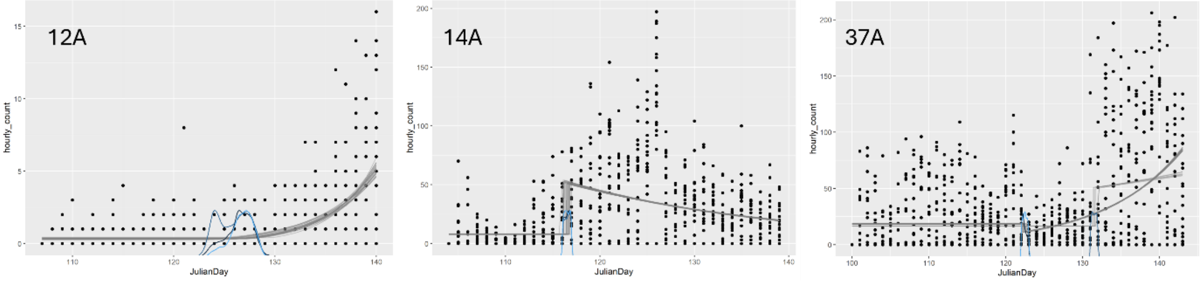

Figure 3.B5, Posterior model fits of hourly vocal detection counts across Julian days for individual nest boxes. The breeding periods included in this analysis are approx. 7 days before the onset of incubation detected to approx. 14 days after. Black points represent observed hourly counts, black lines show posterior mean fit, and blue lines indicate the posterior distribution of estimated changepoints. Each panel corresponds to a different nest box, ID shown in top left corner.

Estimating Changepoints between Incubation and Post-hatching

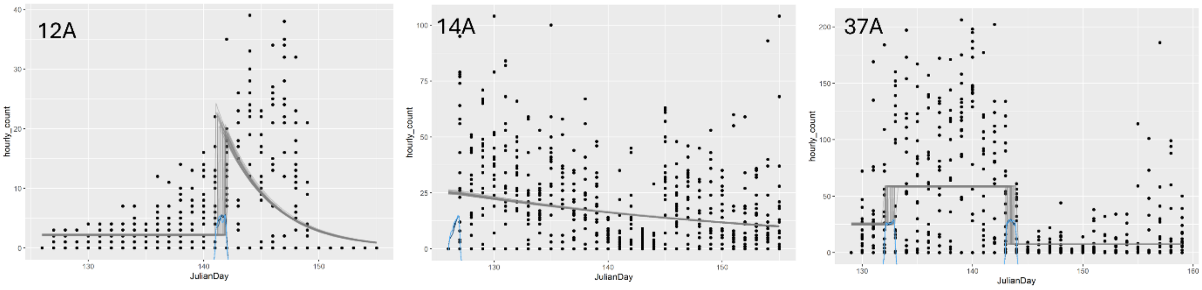

Figure 3.B6, Posterior model fits of hourly vocal detection counts across Julian days for individual nest boxes. The breeding period included in this analysis are approx. 7 days before the onset of hatching to approx. 15 days after, corresponding to F. hypoleuca incubation and post-hatching period. Black points represent observed hourly counts, black lines show posterior mean fit, and blue lines indicate the posterior distribution of estimated changepoints. Each panel corresponds to a different nest box, ID shown in top left corner.

Estimating Changepoints between Post-hatching and Fledging

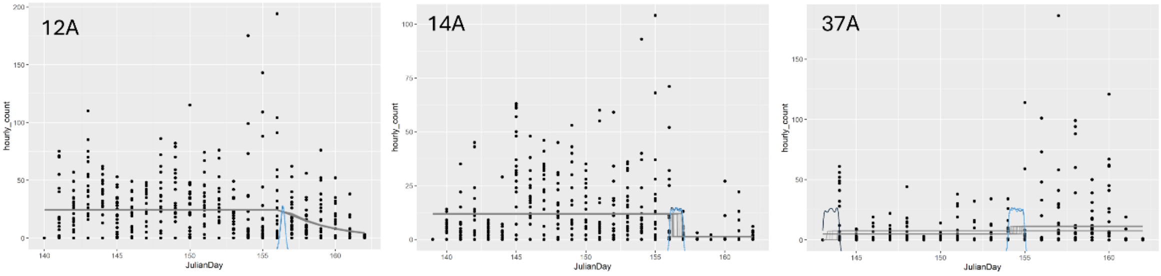

Figure 3.B7, Posterior model fits of hourly vocal detection counts across Julian days for individual nest boxes. The breeding period included in this analysis are approx. 15 days before the onset of fledging to approx. 5 days after, corresponding to F. hypoleuca post-hatching period. Black points represent observed hourly counts, black lines show posterior mean fit, and blue lines indicate the posterior distribution of estimated changepoints. Each panel corresponds to a different nest box, ID shown in top left corner.

3.6.3. Potential for Visitation Rate as a Proxy for Chick Weight Gain

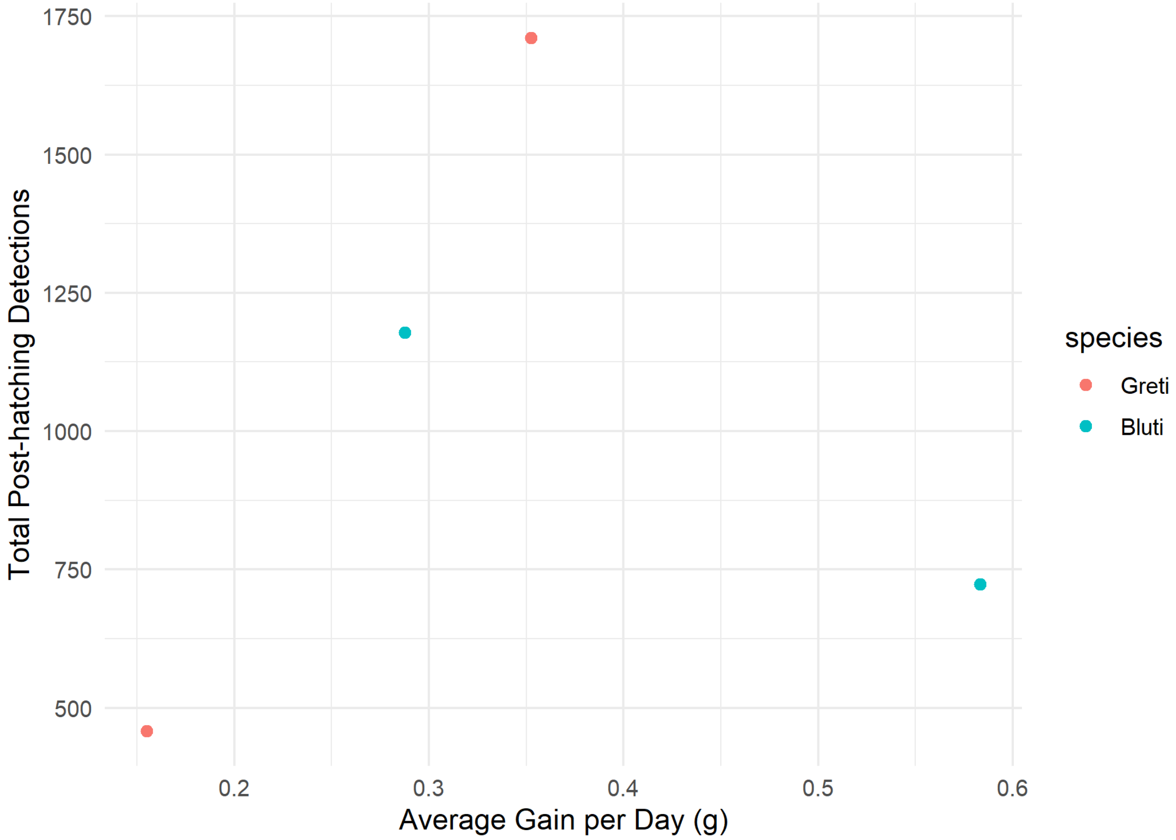

Figure 3.B8, Relationship between total post-hatching detections and nestling growth rate (average gain per day, g) for C. Caeruleus (Bluti) and P. major (Greti). Each point represents a nest box with paired measurements of total detections recorded after hatching and chick growth rates.

4. From Proof of Concept to Application: Linking Breeding Phenology with Migration Modelling

This thesis tested a proof-of-concept for remote nest box monitoring using audio recorders. The aim was to evaluate the feasibility of detecting breeding behaviours and to assess whether changes in frequency of those behaviours could indicate transitions in breeding stage phenology and provide metrics for breeding success.

Audio recorders successfully detected behaviours such as vocalisations, parental visitation, and chick begging. When compared with in-field estimates, both the onset of chick begging and fledging dates were reliably estimated. In the absence of visitation data, vocalisations provided an estimate that captured the interval between male and female arrival dates. However, visitation patterns did not reliably estimate brood size, potentially due to the limited number of nest boxes in this proof of concept. Even so, audio data could be used to generate binary metrics for breeding success, such as brood success or failure.

Beyond insights in breeding status, this thesis highlights important practical considerations. Compared to other remote-sensing techniques such as cameras, audio recorders tend to be less expensive and more flexible to deploy with minimal nest disturbance (Hereward et al., 2021; Prinz et al., 2016; Surmacki & Podkowa, 2022). Measures of breeding behaviour can also be detected without the need for intrusive methods using RFID or PIT tags (Harrison & Kelly, 2022; Nicolaus et al., 2008). Audio recorders therefore offer advantages for scalability. However, recordings require considerable processing time and data storage. Much of this reflects the need for manual validation, which can be partially mitigated by reducing recording schedules and refining classifier performance.

At present, this proof-of-concept remains limited by its small sample size and focus on a single breeding area. Links between visitation rates and breeding success remain unclear. Moreover, disentangling behaviour frequency from detection frequency is challenging without applying more complex models that account for environmental factors influencing both behaviour and flexibility.

To address the remaining challenges, audio recorders should be integrated into existing long-term nest box monitoring networks. Installing recorders under a reduced recording schedule would add minimal fieldwork when compared to constant recording due to fewer battery changes. At the same time, this integration would generate sufficient data to refine classifiers and improve their robustness to variation in the surrounding soundscape, reducing reliance on manual annotation. Combining audio recording with fine-scale measurements of nest box surrounding habitat and microclimate using loggers (Shutt et al., 2022), would help comparisons between nests when estimating brood sizes or drivers of behavioural differences. Expanding the proof of concept would also allow the development and evaluation of a standardised workflow, spanning collection, processing, and analysis, as proposed in Figure 4.1.

The final objective of this thesis aimed to address whether the breeding phenology and success data collected from nest box recordings could inform individual-based migration models. The findings of this thesis support a combined approach. Audio recordings could replace labour-intensive and costly efforts to estimate arrival dates and breeding phenology, while in-person monitoring focuses on measures of breeding success (Figure 4.1). Collectively, these data streams offer a powerful means of generating the empirical information required to parameterise key functions of individual based migration models such as the Spatially-explicit Adaptive Migration Model (Howard et al., 2024).

This contribution would be especially timely. Increasing environmental uncertainty is driving the decline of long-distance migrant populations (Howard et al., 2020). Refining migration models that can both explain and predict migratory behaviour will be essential for understanding population dynamics and informing conservation strategies (Howard et al., 2024).

Identifying audio recorders as a tool to monitor nest-level breeding phenology and success in cavity-nesting passerines represents a novel extension of acoustic behavioral monitoring to estimate breeding status (Bennett et al., 2025; Berger-Tal et al., 2011; Schackwitz et al., 2020; Teixeira et al., 2019, 2024). Applying this methodology to migratory modelling highlights the potential to integrate empirical monitoring with mechanistic modelling to advance predictive capacity and guide conservation action.

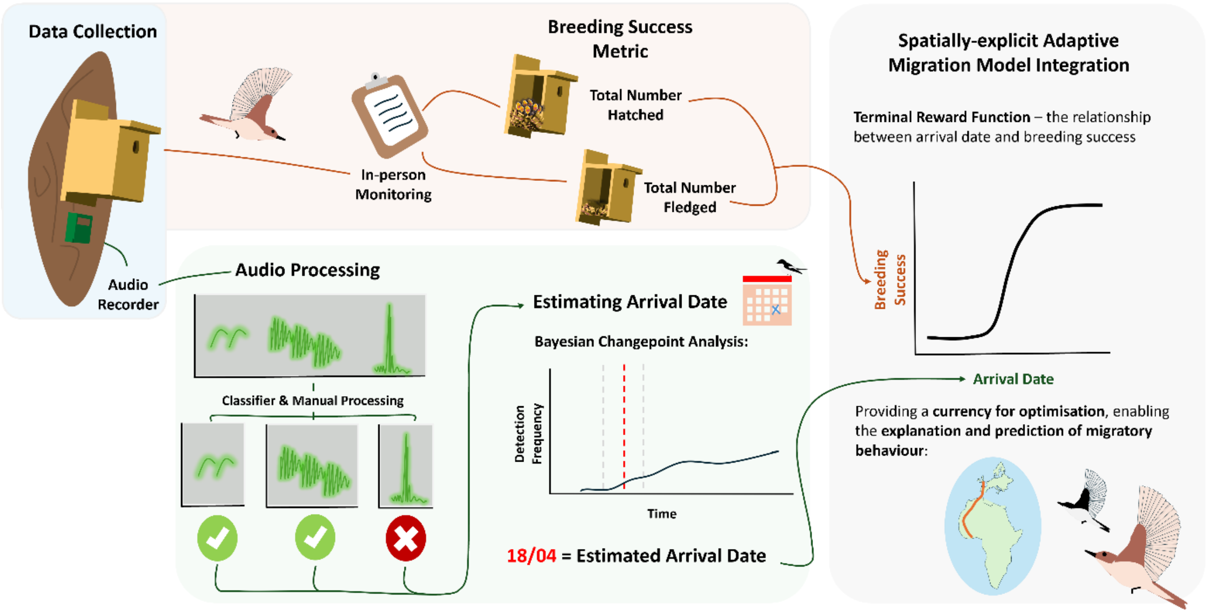

Figure 4.1, A schematic workflow of combining data from remote audio recordings and in-person monitoring to estimate breeding phenology (arrival dates) and metrics for breeding success (brood size) to parameterise the terminal reward function in the Spatially-explicit Adaptive Migration Model.

References

Bennett, S., Williamson, L., Hernández-González, M., Denton, T., Laber, R., Deakin, Z., Bolton, M., Manilow, E., & Wilson, L. J. (2025). Bioacoustics as a Measure of Population Size and Breeding Success of European Storm Petrels Hydrobates pelagicus. *Ecology and Evolution*, *15*(8). https://doi.org/10.1002/ECE3.71893

Berger-Tal, O., Polak, T., Oron, A., Lubin, Y., Kotler, B. P., & Saltz, D. (2011). Integrating animal behavior and conservation biology: A conceptual framework. *Behavioral Ecology*, *22*(2), 236–239. https://doi.org/10.1093/beheco/arq224

Harrison, N. D., & Kelly, E. L. (2022). Affordable RFID loggers for monitoring animal movement, activity, and behaviour. *PLOS ONE*, *17*(10), e0276388. https://doi.org/10.1371/JOURNAL.PONE.0276388

Hereward, H. F. R., Facey, R. J., Sargent, A. J., Roda, S., Couldwell, M. L., Renshaw, E. L., Shaw, K. H., Devlin, J. J., Long, S. E., Porter, B. J., Henderson, J. M., Emmett, C. L., Astbury, L., Maggs, L., Rands, S. A., & Thomas, R. J. (2021). Raspberry Pi nest cameras: An affordable tool for remote behavioral and conservation monitoring of bird nests. *Ecology and Evolution*, *11*(21), 14585–14597. https://doi.org/10.1002/ece3.8127

Howard, C., Mason, T. H. E., Baillie, S. R., Border, J., Hewson, C. M., Houston, A. I., Pearce-Higgins, J. W., Bauer, S., Willis, S. G., & Stephens, P. A. (2024). Explaining and predicting animal migration under global change. *Diversity and Distributions*, *30*(2). https://doi.org/10.1111/ddi.13797

Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., Butchart, S. H. M., & Willis, S. G. (2020). Disentangling the relative roles of climate and land cover change in driving the long-term population trends of European migratory birds. *Diversity and Distributions*, *26*(11), 1442–1455. https://doi.org/10.1111/ddi.13144

Nicolaus, M., Bouwman, K. M., & Dingemanse, N. J. (2008). Effect of PIT Tags on the Survival and Recruitment of Great Tits Parus major.

Https://Doi.Org/10.5253/078.096.0215, *96*(2), 286–292. https://doi.org/10.5253/078.096.0215

Prinz, A. C. B., Taank, V. K., Voegeli, V., & Walters, E. L. (2016). A novel nest-monitoring camera system using a Raspberry Pi micro-computer. *Journal of Field Ornithology*, *87*(4), 427–435. https://doi.org/10.1111/jofo.12182

Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., Woodruff, J., Schackwitz, W., Airola, D. A., Greene, A., Schackwitz, M., & Woodruff, J. (2020). *Journal of Fish and Wildlife Management | www.fwspubs.org*. *11*. https://doi.org/10.3996/102019-JFWM-083

Shutt, J. D., Bell, S. C., Bell, F., Castello, J., El Harouchi, M., & Burgess, M. D. (2022). Territory-level temperature influences breeding phenology and reproductive output in three forest passerine birds. *Oikos*, *2022*(8). https://doi.org/10.1111/oik.09171

Surmacki, A., & Podkowa, P. (2022). The use of trail cameras to monitor species inhabiting artificial nest boxes. *Ecology and Evolution*, *12*(2), e8550. https://doi.org/10.1002/ECE3.8550

Teixeira, D., Maron, M., & van Rensburg, B. J. (2019). Bioacoustic monitoring of animal vocal behavior for conservation. *Conservation Science and Practice*, *1*(8), e72. https://doi.org/10.1111/CSP2.72

Teixeira, D., Roe, P., van Rensburg, B. J., Linke, S., McDonald, P. G., Tucker, D., & Fuller, S. (2024). Effective ecological monitoring using passive acoustic sensors: Recommendations for conservation practitioners. In *Conservation Science and Practice* (Vol. 6, Issue 6). John Wiley and Sons Inc. https://doi.org/10.1111/csp2.13132