Validating a model-mediated acoustic sensor network for biodiversity assessment in Shropshire

Technical evidence report for ornithologists, ecologists, and policy stakeholders.

Harper Adams University

K.J. Mhango, M.J. Lewis, W. Harris. E.

Preface

This document is prepared as a technical evidence report for biodiversity decision-making in the context of the public- interest. Its purpose is to evaluate whether a model-mediated acoustic monitoring network can provide robust, repeatable, and policy-relevant evidence on avian community composition at landscape scale.

The report applies transparent analytical controls, explicit uncertainty boundaries, and a structured interpretation framework. Results are presented as detection-based ecological indicators (presence, recurrence, seasonal structure, and conservation-interest distribution) and not as direct abundance estimates.

The intended use is to support screening, prioritisation, and expert-led follow-up in partner workflows. It should be read as a decision-support evidence base that complements, rather than replaces, field ornithology and formal species identification.

Disclaimer

This assessment draws on coordinated contributions across acoustic sensor deployment, data engineering, and ecological interpretation.

We acknowledge the partners and stakeholders of the River Severn Partnership (RSP) - Advanced Wireless Innovation Regions project, who provided the funding for this study and helped to identify the deployment sites of the acoustic sensor network. We also acknowledge project collaborators and reviewers from Harper Adams University, whose technical feedback improved methodological clarity, reporting precision, and interpretive safeguards for unusual detections, and who provided valuable insights into the ecological context of the study area.

Any remaining interpretation choices and reporting limitations are the responsibility of the authors.

Executive summary

The Shropshire acoustic sensor network generated a strong and coherent community signal over the focal survey window (1 July 2025 to 16 February 2026). On the evidence assembled here, it merits recognition as a credible option for biodiversity assessment when its outputs are interpreted as model-mediated detections rather than direct measures of abundance. The cleaned analytical dataset retained 865,813 detections; the event-collapsed ecological working set contained 107,946 collapsed events across 25 stations and 129 detected species. At network scale, 111 of the 129 detected species were represented in the configured historical GBIF baseline, giving 86.0% overlap between the sensor-derived list and historical context.

Credibility is strongest for community-level presence, richness, broad spatial structure, and repeated seasonal activity. The most frequently detected species - Common Wood-Pigeon, European Robin, Eurasian Jackdaw, Eurasian Wren, Eurasian Blue Tit, and Eurasian Collared-Dove - were all repeatedly detected across most of the station network and showed high event-level confidence distributions. Station richness remained high throughout the network, averaging 69.8 species per station and reaching 92 species at the richest sites.

The system also appears informative for conservation screening. A total of 41 species listed as Red or Amber under the fifth edition of the Birds of Conservation Concern (BOCC5) were detected, including 14 Red-listed taxa. Red-list species richness was concentrated at a small subset of stations (notably PUC-HAU-05, PUC-HAU-21, PUC-HAU-24, and PUC-HAU-16), indicating that the network is sensitive not only to common species but also to conservation-relevant assemblages.

The evidence for rare species is more nuanced. Several historically sparse taxa were repeatedly and spatially replicated - notably Hooded Crow and Red-billed Chough - while other recurrent detections remained historically unsupported and therefore require targeted expert review rather than immediate acceptance. Importantly, the tool is validated here as a credible biodiversity assessment instrument, but not as a replacement for expert adjudication of every unusual record.

Overall, the results indicate that model-mediated acoustic monitoring can recover a biologically plausible and conservation-relevant picture of avian diversity at landscape scale. It is especially valuable for repeated, standardised, high-coverage biodiversity assessment. Its strongest use case is to detect species presence, richness, seasonal structure, migration windows, and conservation-interest assemblages with a level of consistency that makes it suitable for operational monitoring.

Introduction and rationale

This report presents a continuous, designed acoustic biodiversity study in Shropshire using fixed autonomous recording units and AI-assisted species identification. The central rationale is methodological: determine whether model-mediated acoustic detections can provide ecologically credible, repeatable evidence on bird community structure at landscape scale.

The study is motivated by a common monitoring constraint. Manual surveys remain essential but are often limited in temporal density and spatial consistency, especially when evidence is needed repeatedly across many sites. A standardised passive-acoustic workflow offers an alternative by sampling continuously, preserving auditable records, and enabling common analytical rules across all stations and dates.

The objective is therefore not to replace expert ornithology, but to evaluate whether automated detection outputs are robust enough for presence, richness, seasonal timing, and conservation-screening applications when interpreted through transparent quality controls and historical context.

General methods description

This methods synthesis follows a single end-to-end analytical pipeline. Continuous station audio is processed into time-stamped, georeferenced candidate detections; records are then quality controlled and harmonised into analysis-ready tables.

To reduce inflation from repeated model triggers over short intervals, raw detections are collapsed into ecological event units (five-minute windows) and daily station-species summaries. Core interpretation is then based on repeated detection days and station recurrence rather than raw classifier output volume.

In this report, a “collapsed event” means one station-species detection cluster where repeated detections within a five-minute window are treated as one ecological event. A “detection-day” means a species is only counted for a date if there are at least four detections for that species on that date.

This is a deliberate two-stage confirmation rule: 1) event-level de-duplication (5-minute collapse) to avoid overcounting rapid repeat triggers from the same calling bout, and 2) day-level qualification (*>=4* detections/day/species) to retain only species-days with repeated within-day support.

Historical context is incorporated using Global Biodiversity Information Facility (GBIF)-derived species and timing references. Comparison between historical records and fixed acoustic detection is treated as pattern-based (presence structure, seasonal timing, relative rank order) rather than as direct abundance equivalence because historical records and fixed acoustic detections come from different observation processes.

The framework includes explicit plausibility stratification for unusual records: rarity class assignment, temporal anomaly diagnostics, spatial replication checks, and conservation-status overlays. Together, these steps provide a reproducible method for separating strong community-level signal from records that require additional expert review.

For the focal synthesis in this report, the cleaned analytical stream contains 865,813 retained detections and 107,946 collapsed events across 25 active stations, covering 1 July 2025 to 16 February 2026 (Table 1).

Table 1. Core network metrics

Metric	Value
Retained cleaned detections	865,813
Collapsed events used for ecological summaries	107,946
Sensor stations	25
Network species richness	129 species
Historical baseline richness in scope	111 species
Sensor-historical overlap	111 species (86.0% of sensor list)
Sensor-only species	18
Red/Amber species detected	41 (14 Red, 27 Amber)
Mean station richness	69.8 species
Richest stations	PUC-HAU-06, PUC-HAU-05, PUC-HAU-18

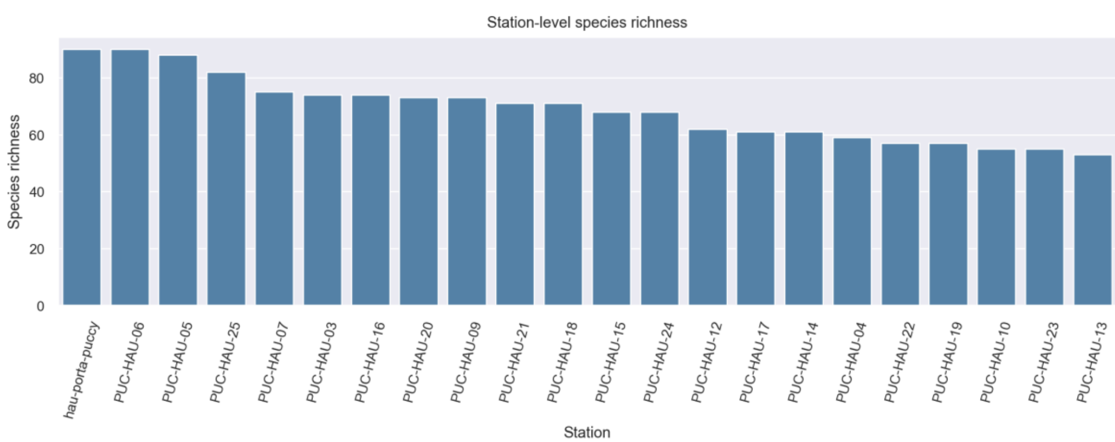

Figure 1. Station-level species richness across July 2025 to February 2026. Richness remained high across the network, ranging from 44 to 92 species per station, with PUC-HAU-06, PUC-HAU-05, and PUC-HAU-18 occupying the top end of the distribution.

What the network detected

The network recovered a recognisably British lowland bird community, dominated by widespread resident and edge-associated taxa but extending well beyond a small common-core fauna. The 12 most frequently detected species were Eurasian Wren (*Troglodytes troglodytes*), Eurasian Jackdaw (*Corvus monedula*), Dunnock (*Prunella modularis*), Common Wood-Pigeon (*Columba palumbus*), Eurasian Blue Tit (*Cyanistes caeruleus*), Long-tailed Tit (*Aegithalos caudatus*), Rook (*Corvus frugilegus*), Eurasian Magpie (*Pica pica*), Redwing (*Turdus iliacus*), Goldcrest (*Regulus regulus*), and Ring-necked Pheasant (*Phasianus colchicus*). Each of these species was detected repeatedly across the full observation window and across almost every station, which is exactly the kind of reproducible signal one would expect from a network credibly tracking common assemblage structure.

Importantly, species richness was not simply inflated by one exceptionally busy recorder. Although a subset of stations (especially PUC-HAU-07, PUC-HAU-05, and PUC-HAU-06) dominated event intensity, richness across the network remained broad and comparatively even. The richest stations all exceeded 70 species, and even the stations with the lowest species richness supported well above 30 species. This pattern indicates that the system is functioning as a network rather than as a single high-output recorder with peripheral appendages. This network-wide temporal pattern is shown in Figure 2, while between-station richness spread is shown in Figure 1. The dominant species composition and recurrence metrics are summarised in Table 2 and visualised in Figure 3.

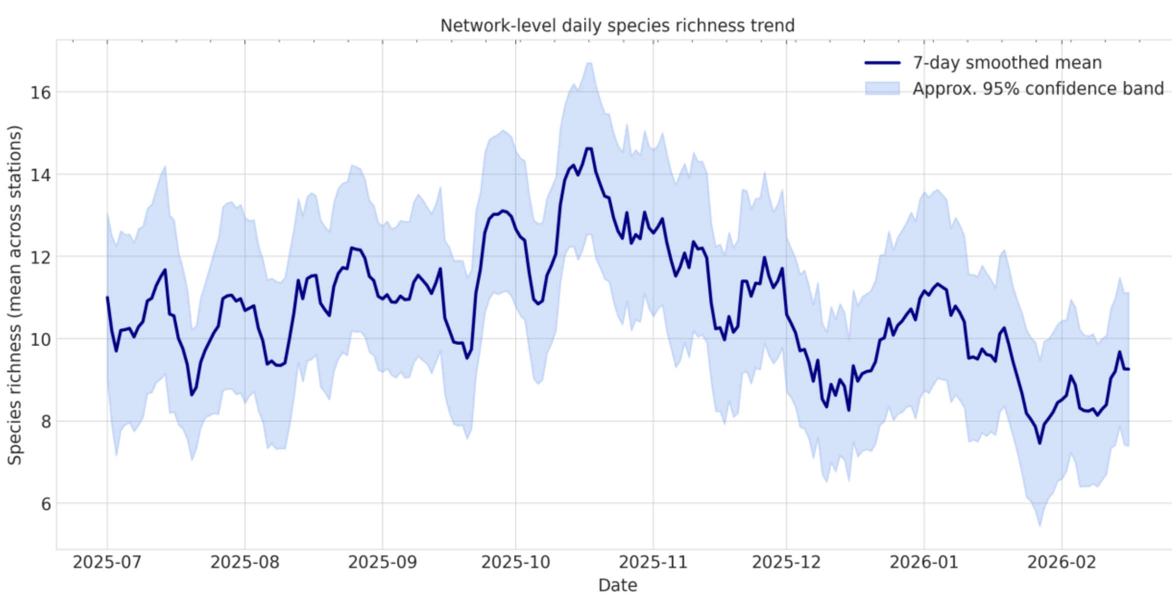

Figure 2. Network-level daily species-richness trend across the observation window, shown as a 7-day smoothed mean with an approximate 95% confidence band across stations.

Table 2. Top detected species by collapsed event count

Common name	**Scientific name**	**Events**	**Detection days**	**Stations**	**Mean event max confidence**
European Robin	*Erithacus rubecula*	11651	231	24	0.89
Eurasian Wren	*Troglodytes troglodytes*	8997	231	25	0.89
Eurasian Jackdaw	*Corvus monedula*	8706	231	22	0.92
Dunnock	*Prunella modularis*	5647	231	25	0.91
Common Wood-Pigeon	*Columba palumbus*	5104	206	25	0.89
Eurasian Blue Tit	*Cyanistes caeruleus*	4687	231	25	0.87
Long-tailed Tit	*Aegithalos caudatus*	4251	223	25	0.91
Rook	*Corvus frugilegus*	4123	228	25	0.87
Eurasian Magpie	*Pica pica*	3909	231	25	0.91
Redwing	*Turdus iliacus*	3659	157	24	0.90
Goldcrest	*Regulus regulus*	3142	227	25	0.86
Ring-necked Pheasant	*Phasianus colchicus*	2970	160	24	0.95

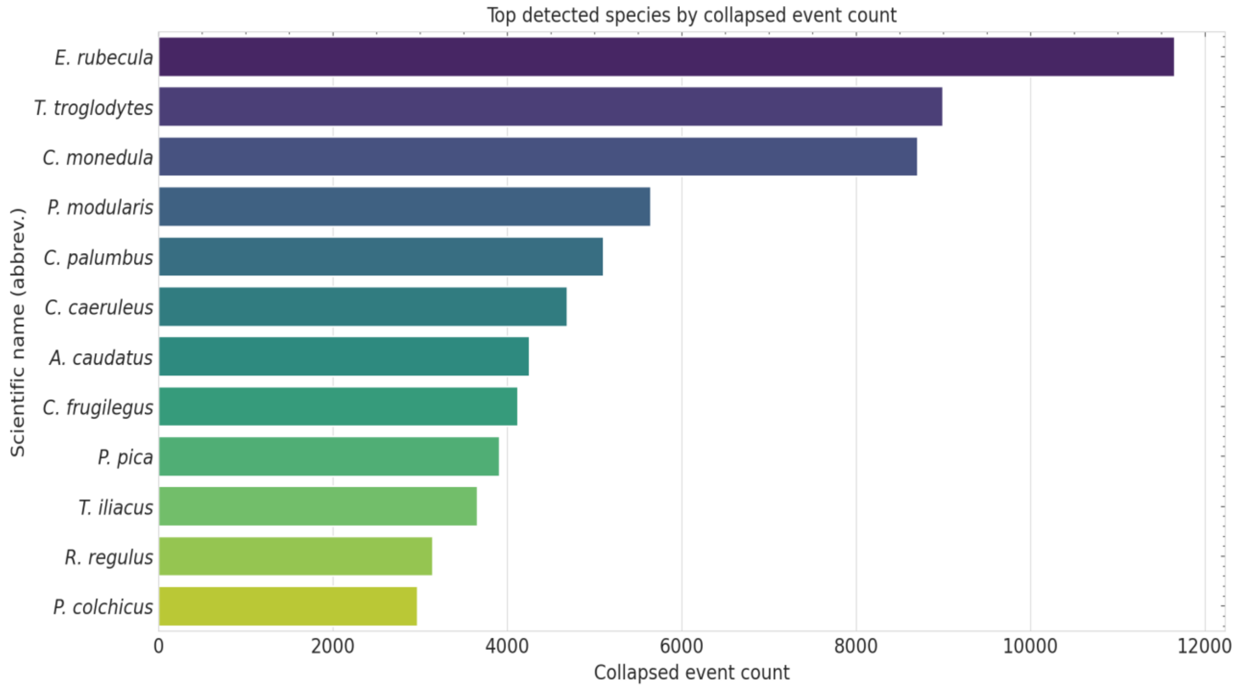

Figure 3. Top species ranked by collapsed event count. The leading species all show broad temporal and spatial recurrence, which is one hallmark of a credible community-level monitoring signal.

How the network compares with historical context

At a county-wide species-list level, agreement with historical context was very high. Of the 129 species detected by the network, 111 were present in the configured historical GBIF baseline. This means that the network broadly recovered the historically expected avifauna for the area, while adding eighteen sensor-only detections that require closer scrutiny.

At the same time, historical comparison should not be over-read. GBIF occurrence records and model-mediated acoustic detections sample birds through different observation processes. One is heterogeneous and historically accumulated; the other is spatially fixed and algorithmically mediated. For that reason, the most defensible comparison is pattern-based rather than count-based. The earlier analytical workflow estimated a sensor-versus-historical Spearman rank correlation of approximately 0.73 and a mean seasonal overlap coefficient of about 0.48, indicating substantial agreement in which species are present but more modest agreement in exactly when activity peaks occur.

The historical-contrast summaries sharpen this interpretation. During the focal breeding-season window, the sensor network recorded 50 species against a historical average of 27.3 in-scope species. The network also recorded 14 Red-list species against a historical average of five, and two summer-migrant warbler species where historical average representation in the configured in-scope window was effectively absent. Those gains should not be interpreted as abrupt ecological change. Rather, they indicate that the acoustic network is a more intensive detection system within the sampled window. The observation window extended through winter, and winter-period species richness and detection-day activity are both above the selected historical average. The full window-by-window comparison is provided in Table 3, and the highest timing deviations are shown in Figure 4.

Table 3. Sensor versus historical comparison within named analytical windows

Analytical window	**Sensor species richness**	**Historical mean richness**	**Richness change**	**Sensor detection days**	**Historical mean days**	**Detection-day change**
Breeding season diversity	50	27.3	83.2%	608	1345.9	-54.8%
Red-list diversity and activity	14	5.0	180.0%	411	368.0	11.7%
Summer migrant warblers diversity and activity	2	0.0	n/a (historical baseline= 0)	2	0.0	n/a (historical baseline= 0)
Winter period diversity and activity	63	22.9	175.1%	2828	1211.1	133.5%

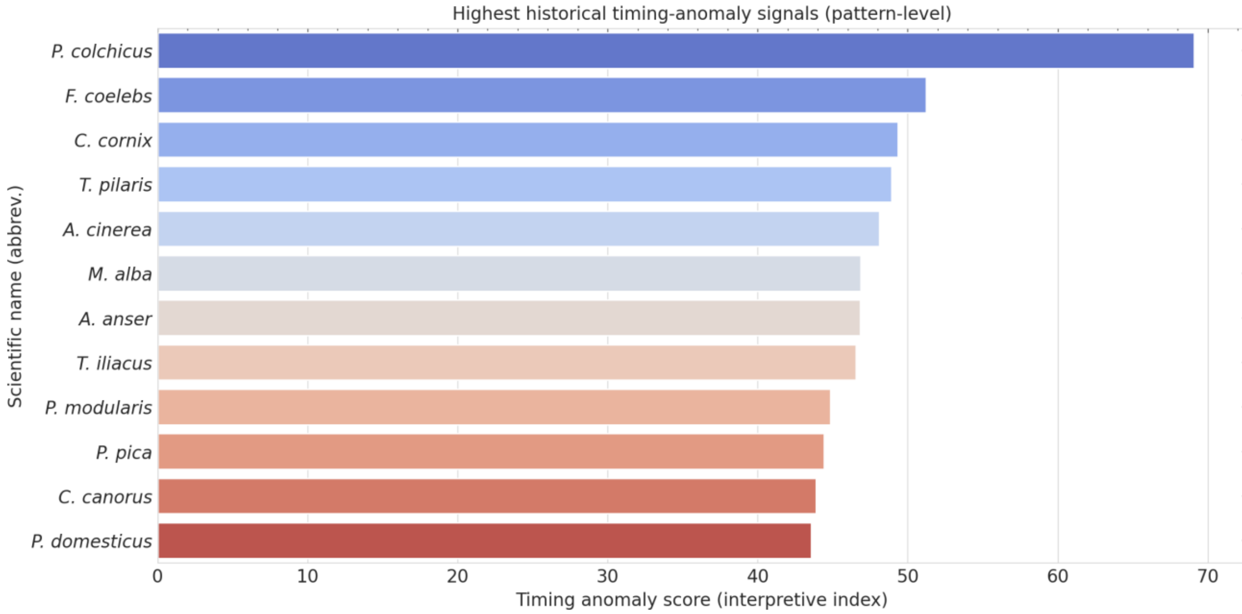

Figure 4. Highest pattern-level historical timing anomalies. These scores identify species whose focal-window timing diverged most strongly from the historical envelope; they are best read as a prioritised review list rather than as direct evidence of ecological change.

Rare, sparse, and unusual detections

In this dataset, rarity was not treated as a binary yes-or-no attribute. Species were partitioned into four classes: *historically expected*, *locally scarce*, *historically sparse*, and *unsupported*. This structure is essential because a detection can be unusual without being implausible, and recurrent unsupported detections can signify either a real boundary case in the baseline or a persistent classification problem.

Important definition note: throughout this report, *`rarity class`* means a historical GBIF support class derived from local *`gbif_hist_count`*, not ecological rarity in the UK. In other words, *`unsupported`* means "not represented in the configured local GBIF window" rather than "biologically rare." European Greenfinch is the clearest example: it is common in UK farmland contexts but appears as *`unsupported`* here due to baseline support mismatch.

The rarity-class distribution was dominated by expected species, but the network nevertheless captured a meaningful tail of high-interest taxa. Twelve species were locally scarce, six were historically sparse, and eighteen were unsupported in the baseline classification. Under the stricter species-day qualification rule used here (a day counts only when a species has at least four detections), only a subset of these remain in the highest-priority context table. The key interpretive question is not merely whether these species appeared, but whether they were repeated in time, spread across stations, and supported by confidence structure consistent with genuine detections.

By that standard, several historically sparse taxa deserve serious attention. Hooded Crow was detected on 86 qualified detection-days across 24 stations, with its strongest local concentration at PUC-HAU-16. Red-billed Chough remained present but at much lower qualified-day frequency, while most sparse taxa fell toward low-frequency tails under the stricter day-count rule. These are still not automatically confirmed records and warrant expert review rather than immediate acceptance.

To highlight the most ecologically interesting unsupported records, the timeline below focuses on the top three unsupported migratory species (excluding expected resident/group-label cases such as Greenfinch and Ring-necked Pheasant). These trajectories are shown in Figure 5.

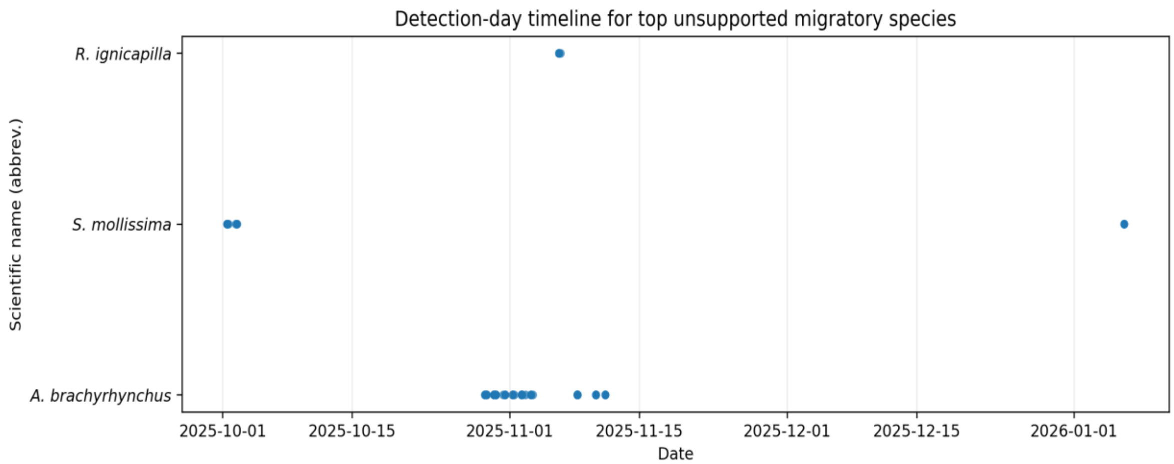

Figure 5. Detection-day timeline for the top three migratory unsupported species (one dot per qualified species-day, where a day is counted only if the species has at least four detections).

The timeline highlights three unsupported migratory detections with very different plausibility profiles:

- **Pink-footed Goose (*Anser brachyrhynchus*)** was detected on 9 qualified detection-days across 4 stations (October-November). This is seasonally plausible for migratory geese, but still requires careful verification because it remains unsupported by the configured local historical frame.
- **Common Eider (*Somateria mollissima*)** persisted at low qualified frequency (3 qualified detection-days at one station). The species is a true sea duck, rarely found away from coasts (*RSPB*). Inland detections are plausible as occasional vagrancy, but they remain high-priority verification targets.
- **Common Firecrest (*Regulus ignicapilla*)** appeared on one qualified detection-day across 4 stations in November. Although not impossible under regional winter movement contexts, this low-frequency unsupported signal is best treated as provisional pending direct audio review.

From a classifier-risk perspective, these unsupported migratory detections are exactly the class where caution is needed. Recent BirdNET guidance emphasizes species-specific thresholding and careful validation of rare/high-impact detections rather than relying on a single universal threshold (*Oecologia/BirdNET threshold guidance*, *detector score interpretation guidelines*). Accordingly, this report treats these findings as *plausible but unresolved* pending targeted expert/audio review and weather-context cross-checking.

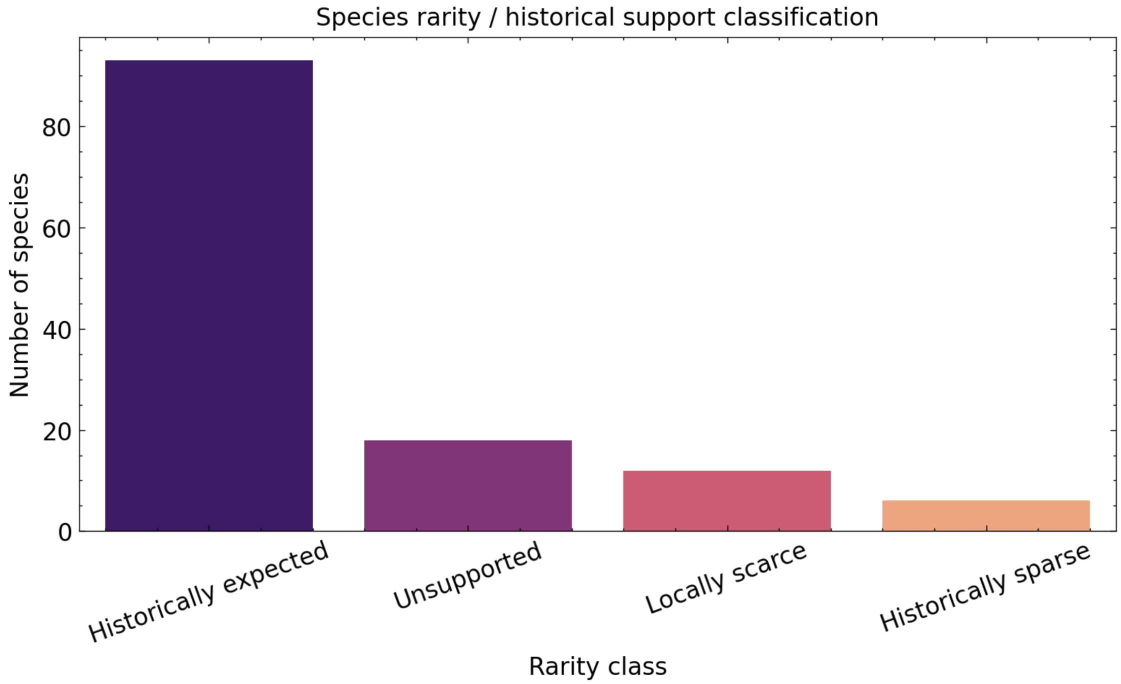

Figure 6. Historical-support classes (not ecological rarity). Most detected species were historically expected, but the network also surfaced a tractable set of locally scarce, historically sparse, and unsupported taxa requiring differentiated interpretation. Species-level context and station-concentration summaries are provided in Table 4 and Table 5.

Table 4. High-interest detections in historical-support context

Common name	**Scientific name**	**Historical-support class**	**Detection days**	**Stations**	**Months**
Hooded Crow	*Corvus cornix*	historically sparse	86	24	7
European Greenfinch	*Chloris chloris*	unsupported	80	23	7
Ring-necked Pheasant	*Phasianus colchicus*	unsupported	65	18	3

Common name	Scientific name	Historical-support class	Detection days	Stations	Months
Pink-footed Goose	*Anser brachyrhynchus*	unsupported	9	4	2
Common Eider	*Somateria mollissima*	unsupported	3	1	2
Red-billed Chough	*Pyrrhocorax pyrrhocorax*	historically sparse	2	7	2
Common Firecrest	*Regulus ignicapilla*	unsupported	1	4	1

Table 5. Where top high-interest detections were concentrated

Common name	Historical-support class	Top station	Days at top station	Peak month	Days in peak month	Stations overall
Hooded Crow	Historically sparse	PUC-HAU-16	38	10	24	24
European Greenfinch	Unsupported	PUC-HAU-05	44	7	21	23
Ring-necked Pheasant	Unsupported	PUC-HAU-15	63	1	29	18
Pink-footed Goose	Unsupported	PUC-HAU-09	9	11	6	4
Common Eider	Unsupported	PUC-HAU-17	3	10	2	1
Red-billed Chough	Historically sparse	PUC-HAU-01	2	7	1	7
Common Firecrest	Unsupported	PUC-HAU-07	1	11	1	4

Conservation value: Red- and Amber-listed species

One of the clearest strengths of the Shropshire network is its ability to recover conservation-relevant assemblages. In total, 41 Red- or Amber-listed species were detected, including 14 Red-listed species. Among the most prevalent Red-listed taxa were Spotted Flycatcher, European Greenfinch, House Sparrow, Yellowhammer, Common House-Martin, Eurasian Curlew, and Fieldfare. Several were detected on dozens of days and across large fractions of the station network.

The distribution of Red-list detections was not uniform. The highest Red-list species richness was recorded at PUC-HAU-05, while PUC-HAU-21 contributed the largest volume of Red-list events. Both species richness and activity also ranked highly at PUC-HAU-24 and PUC-HAU-16 . From an ornithological perspective, this pattern is highly informative. It suggests that the sensor network can do more than compile a species list; it can identify where conservation-sensitive assemblages concentrate and therefore where ground verification, habitat interpretation, or management attention should be focused. The most active Red-listed species are listed in Table 6, and their station-level concentration is shown in Figure 7.

Table 6. Most active Red-listed species in the focal dataset

Common name	**Scientific name**	**Events**	**Detection days**	**Stations**
House Sparrow	*Passer domesticus*	823	107	10
European Greenfinch	*Chloris chloris*	667	80	23
Spotted Flycatcher	*Muscicapa striata*	468	53	20
Fieldfare	*Turdus pilaris*	466	54	19
Yellowhammer	*Emberiza citrinella*	456	43	14
Eurasian Curlew	*Numenius arquata*	440	24	10
Mistle Thrush	*Turdus viscivorus*	229	18	11
Northern Lapwing	*Vanellus vanellus*	211	15	2
Eurasian Tree Sparrow	*Passer montanus*	130	4	2
Common House-Martin	*Delichon urbicum*	111	7	12

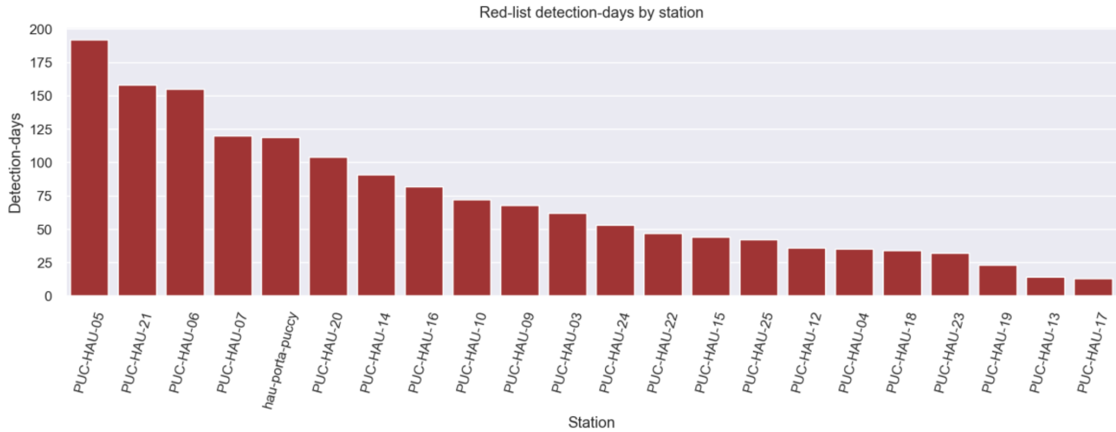

Figure 7. Station-level concentration of Red-list detections. PUC-HAU-05 carried the highest Red-list richness, while PUC-HAU-21 combined high richness with the largest Red-list event volume.

Detection strength, reproducibility, and robustness

The strongest argument for credibility lies not in any single species record but in the overall reproducibility of the signal. One hundred and four of the 129 species detected were recorded on at least five days and at three or more stations. This means that most of the network's output was neither temporally thin nor spatially isolated.

Confidence and probability are distinct model outputs in this workflow: confidence represents detection-level classifier confidence, while probability represents model-assigned likelihood for the species hypothesis. Threshold pairs are therefore reported explicitly.

The prediction confidence of the most common species reinforces that conclusion. The event-level maximum confidence distributions for Wood-Pigeon, Jackdaw, Robin, and Wren are tightly concentrated at high values, while more variable species such as Blue Tit and Collared-Dove still show medians well above the lower threshold bands. These confidence distributions are shown in Figure 8.

Threshold sensitivity analysis provides a second test. When both confidence and probability thresholds were set at 0.3/0.3, the network retained 75 species and 94,468 events across all 25 stations. At 0.5/0.5, it still retained 45 species, and even under the stricter 0.7/0.7 setting it retained 26 species and 57,718 events. In other words, the core community signal survives stricter evidence filters. What drops away under stricter thresholds is mostly the long tail of weaker or less frequent detections, which is the expected behaviour of a well-calibrated biodiversity screen.

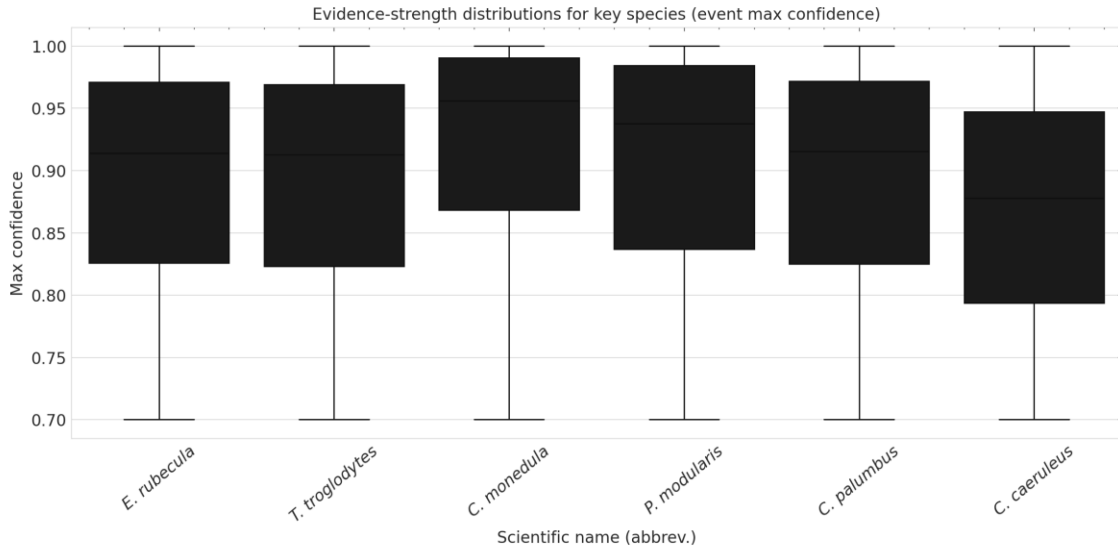

Figure 8. Event-level confidence distributions for key high-frequency species. The tight, high-confidence distributions are consistent with a reproducible common-species signal.

Migration phenology and seasonal windows

The Shropshire network also recovered broad migration windows for a subset of species. These windows should not be interpreted as exact arrival and departure dates, but they do demonstrate that repeated model-mediated detections can recover ecologically interpretable seasonal structure. Species such as Common Swift, Common House-Martin, Barn Swallow, Reed Warbler, Willow Warbler, and Eurasian Curlew show broad species timing envelopes within the full focal study period with identifiable peak positions. This supports the use of the network for phenological surveillance provided that outputs are communicated as windows with uncertainty, not as exact census dates. Estimated windows and peaks are visualised in Figure 9.

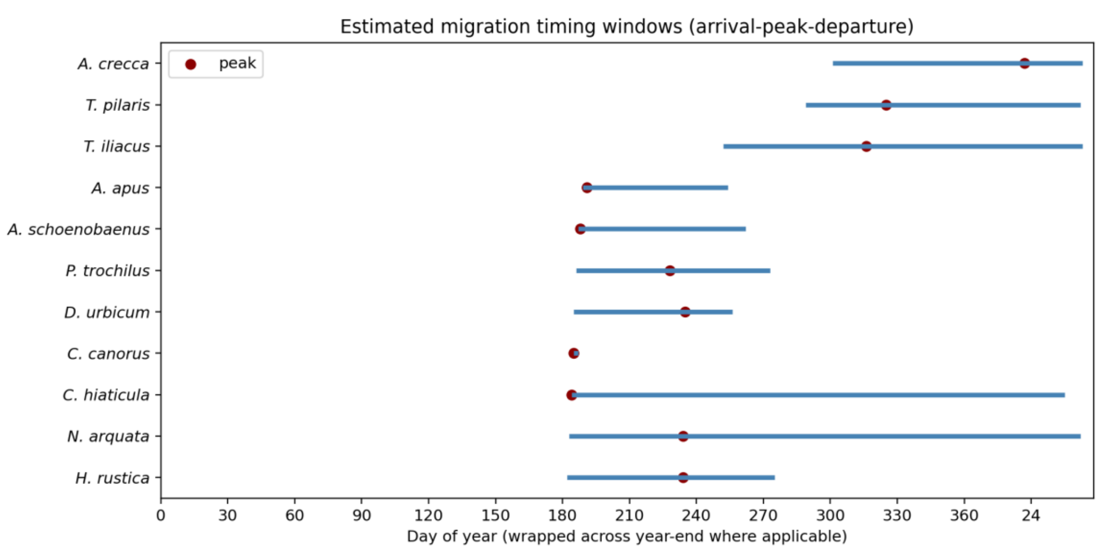

Figure 9. Estimated migration timing windows for selected species. The figure illustrates the network's capacity to recover broad seasonal occupancy windows and internal peak timing from repeated detections.

What the tools can credibly claim - and what they cannot

The present evidence supports three strong claims. First, the acoustic workflow is credible for recovering community-level species presence and richness. Second, it is credible for mapping station-level concentration of activity and conservation-relevant assemblages. Third, it is credible for highlighting repeated sparse taxa and broad seasonal windows in ways that can direct ornithological follow-up.

However, the evidence does not support direct inference about abundance. Model mediated acoustic detections are not currently able to distinguish between individual birds and therefore individual detections could be either separate vocalizations of several individuals or repeated vocalizations of one particularly active individual. Furthermore, the evidence does not support automatic acceptance of every unusual species Historically unsupported or sparse detections still require expert review, habitat or seasonal context, and in some cases direct inspection of source material or classifier behaviour. Despite these points of caution, the value of the network lies in standardised coverage, repeatability, and the capacity to identify biologically plausible structure at scales that would otherwise be difficult to observe consistently.

Closing assessment

Taken as a whole, the Shropshire results validate the acoustic tools used here as a credible option for biodiversity assessment. The network recovered a rich and plausible avifaunal community that aligned strongly with historical presence data, retained a coherent core under stricter evidence thresholds, and detected a substantial portfolio of conservation-priority species. It also demonstrated the ability to distinguish common community structure from a more interpretively delicate tail of sparse and unsupported detections.

For the ornithological community, the key message is that it is now possible to supply an evidence-rich, standardised, and operationally useful layer of biodiversity assessment which when used with clear interpretive safeguards, can materially strengthen landscape-scale bird monitoring.