



## Akustisches Monitoring

### Pilotprojekt 2023 Projektbericht - Endfassung



# Akustisches Monitoring

Pilotprojekt 2023

## Projektbericht - Endfassung

im Auftrag des

Hessischen Landesamt für Naturschutz, Umwelt und Geologie - HLNUG

Staatliche Vogelschutzwarte

Durchgeführt von:

**Dr. Jan O. Engler**

AviCon - Forschung & Planung

Spitzwiesenstr. 32

90765 Fürth

Unter Mitwirkung von:

**Michael Bokämper**

Büro für Studien zur Biodiversität

Reuthstr. 12

91099 Poxdorf

**Stefan Hannabach**

Hartmannstr. 136

91058 Erlangen

Fürth, 18.01.2024

# Zusammenfassung

Das passive ökoakustische Monitoring ist eine moderne und nicht-invasive Methode zur Aufzeichnung hörbarer Biodiversität. Durch Fortschritte in der Aufzeichnungstechnologie und der Geräuscherkennung mit Hilfe von künstlicher Intelligenz hat die Methodik in den vergangenen Jahren enorm an Popularität gewonnen. Jedoch können Einschränkungen wie begrenzte Energieversorgung oder Möglichkeiten zur Fernwartung die langfristige Überwachung der biologischen Vielfalt erschweren, welche eine kontinuierliche Datenerfassung über längere Zeiträume erfordert. Um diese Herausforderungen zu überwinden, haben wir in diesem Pilotprojekt 15 Stationen des hessischen Luftmessnetzes mit speziell angefertigten Audiorekordern ausgestattet, um die Stimmaktivität von Vögeln in städtischen, offenen und bewaldeten Lebensräumen zu erfassen. Die Rekorder sind so eingestellt, dass sie ganztägig die ersten zehn Minuten einer Stunde aufzeichnen und so, zusätzlich zur externen Stromversorgung vor Ort, eine kontinuierliche Datenaufzeichnung über den Untersuchungszeitraum von April bis Oktober 2023 gewährleisten. Durch mobiles Internet werden Audio- und Statusdaten zeitnah zur nachgelagerten Analyse an Cloud-Server versendet. Gleichzeitig ermöglicht dies den Fernzugriff für Service- und mögliche Interventionsmaßnahmen. Im Pilotzeitraum war das System hochgradig zuverlässig mit einer Datenabdeckung von 99,46%. Es konnten 118 Vogelarten aus über 15000 Validierungen bestätigt werden. Für diese Arten wurden "Precision" und Schwellenwerte ermittelt, die es für einen Großteil der bestätigten Arten ermöglichen werden, zukünftig ein (teil-)automatisches Akustikmonitoring in Hessen umzusetzen. Weitere Untersuchungen konnten den Einfluss falsch-positiver Nachweise und die Wahl von Schwellenwerten auf ökologische Aussagen auf Ebene der Artengemeinschaft von Habitattypen und Standorten bis hin zur Einzelart bewerten. Hieraus leiten sich eine Reihe praktischer Empfehlungen für zukünftige Arbeiten ab, die im Bericht ausführlich besprochen werden.

# Summary

Passive eco-acoustic monitoring is a modern and non-invasive method for recording audible biodiversity. Due to advancements in recording technology and sound detection using artificial intelligence, the methodology has gained enormous popularity in recent years. However, limitations such as limited power supply or remote maintenance capabilities can complicate the long-term monitoring of biodiversity, which requires continuous data collection over extended periods of time. To overcome these challenges, we equipped 15 stations of the Hessian air measurement network with specially made audio recorders to capture the vocal activity of birds in urban, open, and forested habitats. The recorders are set to record the first ten minutes of every hour throughout the day, thus ensuring continuous data recording from April to October 2023, in addition to the external power supply on site. Through mobile internet, audio and status data are promptly sent to cloud servers for subsequent analysis. This also enables remote access for service and possible intervention measures. During the pilot period, the system was highly reliable with a data coverage of 99.46%. 118 bird species were confirmed from over 15,000 validations. Precision and threshold values were determined for these species, which will enable (partially) automated acoustic monitoring in Hesse for a majority of the confirmed species in the future. Through further analyses, the Pilot study allowed the assessment of the impact of false-positive detections and the choice of threshold values on ecological signals at the level of species communities across habitat types and locations up to the individual species. From this, a series of practical recommendations for future work are derived, which are discussed in detail in the following report.

# Einleitung

Biodiversitätsmonitoring mit neuen Technologien ist entscheidend für den Umweltschutz. Fortschrittliche Technologien wie Fernerkundung, stationäre Sensorphalanxen, Drohnenüberwachung und künstliche Intelligenz ermöglichen eine präzisere und umfassendere Erfassung von Artenvielfalt und Ökosystemzuständen (Pettorelli et al. 2014, Robinson et al. 2022, Wägele et al. 2022). Diese Methoden liefern wertvolle Daten in Echtzeit und über große, oft schwer zugängliche Gebiete, was zu einem umfassenderen Kenntnisstand und somit effektiveren Schutzmaßnahmen führt (Darras et al. 2019, Wägele et al. 2022). Die Analyse großer Datenmengen erlaubt es, Veränderungen in der Biodiversität schnell zu erkennen und darauf zu reagieren. Dies ist für den Artenschutz und im Naturschutzmanagement unerlässlich. Somit spielt das Biodiversitätsmonitoring eine Schlüsselrolle nicht nur in der ökologischen Forschung und dem Artenschutz, sondern auch in der nachhaltigen Bewirtschaftung natürlicher Ressourcen sowie der Umweltplanung.

Das Aufkommen des Passiven Akustischen Monitorings (PAM) etabliert sich zusehends als eine wichtige Methode im Bereich des Arten- und Biodiversitätsschutzes (z.B. Darras et al. 2019). PAM verwendet hochentwickelte, z.T. kostengünstige Aufzeichnungsgeräte, um die Geräusche in natürlichen Lebensräumen zu erfassen um ein tieferes Verständnis der Anwesenheit, des Verhaltens und der Interaktionen lautgebender Tierarten zu ermöglichen (z.B. Hill et al. 2018, Sethi et al. 2018, Höchst et al. 2022). Diese Technik ist besonders wertvoll für die Überwachung schwer erfassbarer Arten, wie dämmerungs- und nachtaktiver Tiere oder in Schutzgebieten, wo menschliche Störungen zu minimieren sind. Die akustischen Daten ermöglichen es, die Anwesenheit von lautgebenden Arten störungsfrei und über lange Zeiträume hinweg zu dokumentieren und so beispielsweise Veränderungen in der Tierwelt zu erkennen, die auf Umweltveränderungen oder anthropogene Stressoren hinweisen könnten.

Darüber hinaus erleichtert PAM das Langzeitmonitoring von Biodiversität, indem kontinuierlich Daten gesammelt werden, ohne dass eine ständige menschliche Anwesenheit erforderlich ist. Dies reduziert Störungen in empfindlichen Ökosystemen und liefert unverzerrte Daten über die natürlichen Aktivitäten der Tierwelt. Der Einsatz von PAM kann daher in vielen Situationen kostengünstiger und mit einer besseren Datenqualität gegenüber klassischer Erfassungsmethoden eingesetzt werden (Darras et al. 2019, Markova-Nenova et al. 2023) beziehungsweise erlaubt ganz neue Erfassungsansätze die mit klassischen Kartiererfassungen nicht möglich gewesen wären. Die Signalprozessierung mittels künstlicher Intelligenz und maschinellem Lernen ermöglicht erst den umfänglichen Einsatz von PAM, da so eine Analyse der großen Datenmengen realisiert werden kann (z.B. Wägele et al. 2022, Kahl et al. 2021).

Autonome Aufnahmegeräte mit langer Standzeit („autonomous recording units“ „ARUs“, oder auch „Horchboxen“) sind seit etlichen Jahren kommerziell verfügbar und gewinnen spätestens mit der Entwicklung von besonders günstigen Geräten zunehmend an Beliebtheit auch bei der Vogelerfassung (z.B. Hill et al. 2018). Die Auswertung der Tonaufnahmen auf vorkommende Vogelarten ist ohne eine Automatisierung jedoch oftmals zeitraubend und wenig effektiv, bzw. nur für bestimmte Zielarten effizient. Die automatisierte Suche nach Vogelstimmen in den Tonaufnahmen wird jedoch seit wenigen Jahren durch das Aufkommen neuartiger KI-gestützter Algorithmen enorm verbessert. Der bekannteste und am weitesten verbreitete Vertreter dieser neuartigen Erkennungssoftware ist BirdNET (Kahl et al. 2021). Es handelt sich hierbei um frei nutzbare und verfügbare Software, welche die Gesänge und Rufe von weltweit mehreren tausend Vogelarten in Tonaufnahmen erkennen kann. Die breite Verfügbarkeit relativ günstiger Aufnahmegeräte in Kombination mit einer (halb-)automatisierten, Software-gestützten Auswertung macht das PAM für Vogelstimmen zu einem Thema, zu dem aktuell national und international intensiv geforscht und entwickelt wird. Ein kritisches Thema ist dabei die Reduzierung der Fehlerraten der Erkennungsalgorithmen, damit sich der Aufwand für Validierungen reduziert und die Methode in der Praxis verlässlicher und anwendbarer wird (z.B. Perez-Granados 2023).

Die Herausforderungen im Langzeitmonitoring von Vögeln mittels passiver akustischer Überwachungssysteme liegen sowohl in der Hardware der Aufnahmegeräte als auch in der Software zur Datenanalyse. Bei der Hardware ist die Hauptforderung die Entwicklung robuster, energieeffizienter Aufzeichnungsgeräte, die extremen Wetterbedingungen und abgelegenen Standorten standhalten können. Sie müssen über eine lange Laufzeit verfügen, um kontinuierliche Aufnahmen ohne häufige Wartung oder Batteriewechsel zu gewährleisten. Eine weitere Herausforderung besteht darin, Mikrofone und Aufzeichnungssysteme so zu konstruieren, dass sie Aufnahmen mit hoher Klarheit aufzeichnen können, ohne dass Hintergrundgeräusche die Arterkennung beeinträchtigen.

Auf der Softwareseite erfordert die Verarbeitung und Analyse der riesigen Datenmengen, die durch kontinuierliche Aufzeichnungen generiert werden, fortschrittliche KI-Algorithmen und eine fundierte Nachbereitung der so erfolgten Detektionen. Die Herausforderung besteht darin, die Vogelgesänge korrekt zu identifizieren und von anderen Umgebungsgeräuschen unterscheiden zu können. Außerdem ist es wichtig, dass ein Analysesystem in der Lage ist, Veränderungen der Klanglandschaft im Zeitverlauf zu erkennen, um standortspezifische Art-Umweltbeziehungen bewerten zu können. Diese Herausforderungen bedeuten, dass eine ständige Weiterentwicklung und Anpassung der Systeme erforderlich ist, um effektive Langzeitmonitoring-Programme für Vögel zu gewährleisten.

In dieser Pilotstudie soll ein System entwickelt und zum Einsatz gebracht werden, welches die Machbarkeit eines Dauermonitoring für Vögel untersuchen soll. Insbesondere stehen dabei folgende Leitfragen im Fokus:

- **Wie hoch ist die Zuverlässigkeit des Systems?** Welche technischen Hürden treten auf und wie wirken sich diese auf die Laufzeit/Zuverlässigkeit des Systems aus? Wie hoch ist der Wartungsaufwand, der aus dem Betrieb des Systems über lange Zeiträume resultiert?
- **Wie gut ist die Arterkennung?** Wie gut lassen sich verfügbare KI Modelle nutzen und welche Erkenntnisse zur Vogelwelt lassen sich daraus ableiten? Wie gut werden welche Arten erkannt? Wie oft kommt es zu Fehlbestimmungen?
- **Wie lässt sich ein vollautomatisches Dauermonitoring umsetzen?** Für welche Arten eignet sich solch ein System? Wie groß ist die Fehlertoleranz, um ökologische Aussagen nicht zu verzerren? Welche technischen Zwischenschritte sind neben Recorder und KI-basierter Auswertung nötig, um Prozesse zu automatisieren?

Zu diesem Zweck wurden 15 Stationen des Hessischen Luftmessnetzes mit einem eigens entwickelten Horchbox-System auf Basis des ecoSNOOPER-Recorders über einen Zeitraum von über sechs Monaten eingesetzt. Parallel wurden Prozessabläufe entwickelt, die eine zeitnahe Fehlererkennung an den Stationen ermöglichen, eingehende Tondaten mittels KI auswerten und systematisch durch Experten validiert werden. Der nachfolgende Bericht geht detailliert auf die Konzeption des Systems, die durchgeführten Analysen und die Ergebnisse ein und diskutiert diese abschließend im Hinblick auf Handlungsempfehlungen für zukünftige Projekte dieser Art.

# Methoden

## Standorte

Das Hessische Landesamt für Naturschutz, Umwelt und Geologie - HLNUG betreibt ein landesweit ausgerichtetes Messnetz zur gebietsbezogenen Überwachung und Beurteilung der Luftqualität. Dieses Messnetz umfasst insgesamt 38 Stationen, welche die Luftqualität in den Ballungsräumen Rhein-Main und Kassel sowie den Gebieten Lahn-Dill, Nord- und Mittelhessen sowie Südhessen abdecken. Schwerpunkte bilden städtische und verkehrsreiche Gebiete. Diese werden mit ländlichen Standorten, die sich sowohl auf den Wald als auch aufs (Agrar-) Offenland verteilen, verglichen. Für die Pilotstudie wurde angestrebt, urbane Gebiete, Offenlandschaften und Wälder gleichmäßig zu untersuchen.

In Zusammenarbeit mit dem Dezernat I2 "Luftreinhaltung: Immissionen" des HLNUG wurde eine Vorauswahl von Stationen vorgenommen, von denen letztlich 15 Stationen zur Anbringung der Audiorecorder ausgewählt wurden. Vier Stationen liegen in urbanen Räumen, fünf in Offenlandschaften sowie sechs in Waldhabitaten (Abb. 1, Tab. S1). Der größte Vorteil der Nutzung dieser Stationen liegt in der verfügbaren Stromversorgung und in der sicheren Erreichbarkeit während und nach der Installation.

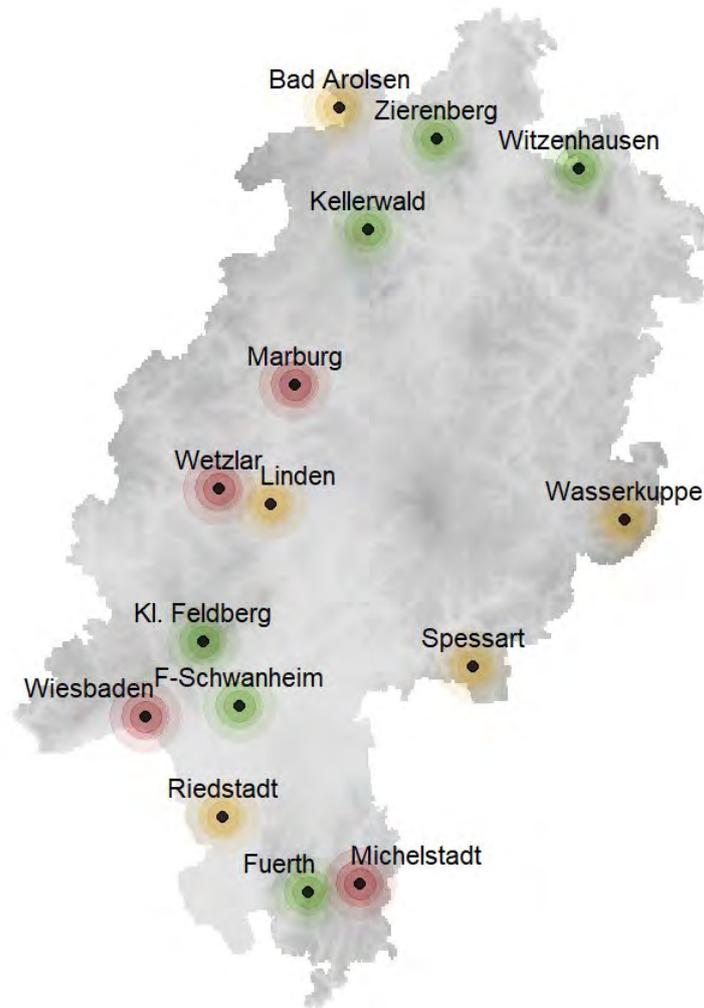


Abb. 1: Übersichtskarte der Standorte der installierten Stationen in Hessen. Standorte urbaner Lagen sind rot dargestellt. Wald- und Offenlandstandorte sind grün bzw. gelb hervorgehoben.

*Fig. 1: Overview map of the locations of the installed stations. Urban location sites are shown in red. Forest and open land locations are highlighted in green and yellow, respectively.*

## Aufnahmegeräte

Die hier zum Einsatz kommenden Recorder sind eine technische Eigenentwicklung. Das Herzstück stellt der ecoSNOOPER dar, ein von der Software- und Hardwareseite her speziell für das Akustikmonitoring entwickeltes Outdoor-Smartphone auf Androidbasis. So kommen u.a. ein spezielles Mikrofon und eine wetterfeste Einhausung zum Einsatz, in welchem sich eine separate Laderegulierung zur Stromversorgung befindet. Im Backend kommen ein 2TB Cloud-Datenspeicher sowie ein lokales NAS Backup System zum Einsatz.

## ecoSNOOPER

Der ecoSNOOPER ist eine Eigenentwicklung von Stefan Hannabach, Michael Bokämper und Jan Engler. Die Basisversion ist seit 2020 in der Entwicklung und im regelmäßigen Einsatz für gezielte Vogelerfassungen v.a. im Kontext von Kartierungen. Das Smartphone bietet eine erprobte und energieeffiziente Hardwarelösung und läuft im Entwicklermodus auf Android 12. Eigens entwickelte Hintergrundprozesse und die grafische Nutzeroberfläche des ecoSNOOPER erlauben, wie bei vielen ARUs, die Einstellung von Aufnahmezeiten und -qualität. Für das Pilotprojekt wurde die bisherige Basisversion um eine Intervallaufnahmefunktion erweitert. Besonders hervorzuheben ist dabei die Rückmeldung via Mobilfunk über durchgeführte Aufnahmen, Speicherplatz und Akkustand des Recorders, welche auch eine Neukonfiguration aus der Ferne via Mobilfunk ermöglicht. Für die Datenübertragung der Aufnahmen selbst wurden mobile Datenpakete der drei gängigen Netzprovider (D1, D2 und E Netz) mit mindestens 15GB je Gerät und Monat gebucht. Neben der Datenübertragung werden die Aufnahmen zusätzlich lokal auf den internen Speicher von 32GB gelegt, der wiederum durch eine 128GB microSD-Karte für die lokale Datensicherung erweitert wurde.

Für das Projekt wurde zu jeder vollen Stunde eine 10-minütige Aufnahme vorgenommen und direkt im Anschluss an den Cloudspeicher übertragen. Aufnahmen wurden im MP3-Format mit einer Bitrate von 192kbs angefertigt. Eigene Vorstudien haben gezeigt, dass bei dieser Komprimierung der Verlust von ökoakustischen Informationen minimal ist im Vergleich zu unkomprimierten Aufnahmen im WAV-Format und nur 25% des Speicherbedarfs benötigt werden. Einmal täglich überprüft jedes Gerät seinen Standort und fragt mögliche neue Konfigurationen ab.

## Gehäuse und Anbringung

Für den Dauerbetrieb wurde der ecoSNOOPER in ein ABS-Wandgehäuse mit IP65 Schutzstandard installiert. In das Wandgehäuse wurden auf der Unterseite zwei Kabeldurchlässe nach IP68 für die Stromzufuhr und das externe Mikrofon gelegt. Für den Anschluss an die Außensteckdose der Luftmessnetzstation wurde in das Wandgehäuse eine 230V Schuko-Steckdose installiert und eine 2,5 Meter lange Zuleitung durch den Kabeldurchlass nach außen geführt. Auf diese Steckdose wurde eine digitale Zeitschaltuhr gesetzt und mit einem nächtlichen Ladeintervall versehen, das dem Stromverbrauch des Smartphones entsprach. Damit wurde einerseits eine Dauerladung verhindert, die dem Akku schaden kann und andererseits die mit dem Ladevorgang einhergehende Wärmeentwicklung auf die kühlen Nachtstunden gelegt. Das Wandgehäuse wurde mit 1-Zoll Manschetten an das Außengeländer der Luftmessnetzstation angebracht und so ausgerichtet, dass Störgeräusche

der Station minimiert wurden. Von außen wurde das Gehäuse mit Warn- und Infoschildern gekennzeichnet, um auf die stattfindenden Aufnahmen hinzuweisen.

## Externes Mikrofon

Das externe Mikrofon ist ebenfalls eine Eigenentwicklung und nutzt eine empfindliche Kondensator-Mikrofonkapsel mit omnidirektioneller Richtcharakteristik (Hersteller PUI Audio, Typ "AOM 5024L"; Signal-Rausch-Abstand: 80dB; Empfindlichkeit: -24dB re 1V/Pa (63mV/Pa)). Für den Einsatz und die Installation am Wandgehäuse wurde ein etwa 15 cm langes Anschlusskabel verbaut. Der Wetterschutz ist konstruktiv, mit einem trichterförmigen Kranz um die Mikrofonbasis zum kontrollierten Abfluss von Regenwasser. Zur Reduzierung von Windgeräuschen wurde ein Schaumstoffschutz um die Kapsel gestülpt. Das Mikrofon nutzt zur Stromversorgung den Kopfhöreranschluss des ecoSNOOPERS („Plug-in-Power“).

## Backend

Der Datentransfer läuft über Google Drive mit einem Zielordner für jedes Gerät. In diesen Ordnern liegen jeweils Unterordner, u.a. für Audiodaten und Statusberichte, sowie Konfigurationsordner für das Ablegen neuer Einstellungen an das Gerät. Der Cloudspeicher wird dabei von einem lokalen Mini PC täglich auf neue Daten hin überprüft. Angaben zu neuen Audiodaten, Akkuständen und Betriebstemperaturen werden in Log-Dateien geschrieben und für die Geräteüberwachung und das in Entwicklung befindliche Frühwarnsystem bereitgestellt. Damit soll ein Überblick über aktuelle Gerätezustände erleichtert und auftretende Probleme frühzeitig erkannt werden. Von der Cloud findet eine physische Datensicherung sowohl am Gerät (auf die externe Speicherkarte) als auch auf einem lokalen NAS Speicher statt.

Für die direkte Gerätefernsteuerung (zum lokalen Archivieren von Daten oder zum Beheben von eventuellen Software-Problemen) kommt eine MDM-Software zum Einsatz. Diese erlaubt das Spiegeln der Benutzeroberfläche auf einen Computer oder ein mobiles Endgerät und somit das direkte Eingreifen in interne Geräteabläufe.

## KI-gestützte Auswertung

Für die automatisierte Bestimmung von Vogellauten nutzen wir BirdNET-Version 2.3. BirdNET, ebenso wie auch andere KI-Algorithmen, erkennt die Vogelstimmen in den Tonaufnahmen anhand einer Mustererkennung und Klassifizierung von normalisierten Spektrogrammen – letztlich also eine Bilderkennung. Die Software gibt bei jeder Erkennung und Klassifizierung immer eine „confidence“ an. Diese dient als Qualitätsmaß wie „gut“ oder „zuverlässig“ das Spektrogramm der Fundstelle den zuvor erlernten und hinterlegten Mustern für eine bestimmte Vogelart entspricht. BirdNET teilt die Aufnahme in drei-Sekunden-

Abschnitte, die jeweils einzeln und unabhängig voneinander analysiert werden. Für die Parametereinstellungen setzten wir die Sensitivität auf 1,4 und die Konfidenz auf 0,3. Die Sensitivität steuert die Empfindlichkeit der Klassifizierung auf leise Signale (bei Werten  $<1$  wird die Analyse auf laute und deutliche Signale beschränkt; bei höheren Werten  $>1$  werden auch leisere Signale einbezogen). Die Einstellung der Konfidenz gibt an, welcher Wert für eine Fundstelle mindestens erreicht werden muss, damit diese Fundstelle in das Ergebnis ausgegeben wird. Hier ist eine entsprechende Filterung auch nach der Ausgabe der Ergebnisse immer noch einfach möglich, bei der Sensitivität ist das nicht der Fall.

BirdNET erlaubt es, die ursprüngliche Anzahl trainierter Arten auf verschiedene Weisen zu filtern. BirdNET 2.3 erkennt über 3200 Vogelarten. Ohne eine geografische Vorfilterung würden sämtliche Arten in die Analyse einbezogen werden. Daher wurden alle Analysen mit einer eigenen Artenliste durchgeführt, die das „Suchschema“ von BirdNET einschränkte. Diese Liste umfasste alle Arten, die in Hessen entweder als Brut-, oder als Rastvogel bekannt sind (Datengrundlage HLNUG).

## Validierung

Die Validierung erfolgte in zwei Stufen. In der ersten Stufe wurden über 15.000 BirdNet Detektionen über das gesamte detektierte Artenspektrum des Zeitraums 15.4.2023 bis 31.7.2023 nach einer gewichteten Zufallsauswahl selektiert und von drei versierten Ornithologen validiert. In einer zweiten Stufe wurden dann weitere 14.000 BirdNet Detektionen überprüft, die a) im selben 3-Sek Zeitfenster detektiert wurden wie die Detektionen aus Stufe 1 oder in dem davor- bzw dahinterliegenden 3-Sek Zeitfenster.

Die Validierungen aus Stufe 1 sollten dazu dienen die Anwesenheit einer Art im Rahmen der Pilotstudie mit hoher Sicherheit zu bestätigen oder auszuschließen. Gleichzeitig sollten für Arten mit bestätigter Anwesenheit Schwellenwerte ermittelt werden, die eine (Halb-)automatische Prozessierung der Nachweise erlauben soll. Als Grundlage wurden aus allen BirdNET Detektionen für jede detektierte Art, jeden Standort und Tag die Detektion mit dem maximalen Konfidenzwert ausgewählt (Tagesmaximalkonfidenz). Aufgrund der Vielzahl von Detektionen und Arten soll hierdurch eine Optimierung der Nachweisbestätigung erreicht werden. Aus der Gesamtheit dieser Tagesmaximalkonfidenzen je Art und Standort wurde eine Gewichtung durchgeführt, nach der selten detektierte Arten einen höheren relativen Anteil an Validierungen zugeteilt bekamen als häufige Arten. Auf diese Weise konnten sämtliche Arten mit einer optimalen Zahl von Validierungen berücksichtigt werden, ohne von häufigen Arten dominiert zu werden. Gleichzeitig ließen sich die bestätigten Nachweise als gesicherte Nachweistage übersetzen. Die errechnete Zahl von Validierungen pro Art war die Grundlage

für die Auswahl an Detektionen. Diese wurden mit absteigendem Konfidenzwert des Nachweises gewählt, um zu verhindern, dass hohe Konfidenzen bei einer Zufallsauswahl nicht gezogen würden.

Neben den beschriebenen Vorteilen dieses Vorgehens ist der Nachteil einer unausgewogenen Validierung der Nachweise, dass Detektionen mit niedriger Konfidenz systematisch unterrepräsentiert werden. Aufgrund der begrenzten Kapazitäten in der Validierung durch die Vielzahl an Detektionen und dem Ziel, akkurate und zuverlässige Obergrenzen zu ermitteln, wurde dieses Vorgehen gegenüber einer gleichmäßig stratifizierten Stichprobe vorgezogen. Die auf diese Weise erzeugte zu validierende Stichprobe wurde schließlich mit "richtig", "falsch" oder "unklar" gekennzeichnet. Unklare Fälle wurden von mehreren Mitgliedern überprüft, bis diese entweder eine klare Zuordnung bekamen oder für die weitere Analyse verworfen wurden. Letztlich wurden 15.033 Detektionen eindeutig als "richtig" oder "falsch" klassifiziert.

Die Validierungen der zweiten Stufe hatten zum Ziel "Begleitarten" (also von BirdNET detektierte Arten zum Zeitpunkt einer Referenzdetektion) der in Stufe 1 bestätigten bzw. widerlegten BirdNET Detektionen zu untersuchen. Hierbei wurde vor allem überprüft, ob eine Echt-Positiv-Detektion von BirdNET in Stufe 2 eine Falsch-Positiv-Detektion aus Stufe 1 erklärt. Dafür wurden sämtliche BirdNET Detektionen (ab einer Konfidenz von 0.3) im selben Zeitintervall der Detektion aus Stufe 1, sowie aus den drei Sekunden vor und nach der Detektion berücksichtigt. Nach diesen Kriterien ergaben sich etwa 14.000 weitere Detektionen von 229 Arten, die nach demselben Schema aktuell noch validiert werden. Eine Gewichtung der Anzahl an Detektionen (wie in Stufe 1) fand nicht statt, da sämtliche Detektionen aus dem betreffenden Zeitraum untersucht werden sollten.

Die Validierung der BirdNET Detektionen erfolgte mit der Software Kaleidoscope Pro (Hersteller: Wildlife Acoustics). Die Vorauswahl wurde mit eigens programmierten R-Skripten realisiert. Gelang eine Validierung nicht innerhalb des drei Sekunden Zeitraums wurde mit der "Reload" Funktion von Kaleidoscope auch der umliegende Zeitraum der Aufnahme überprüft.

## Analyse

Auf Basis der vorgenommenen Validierungen der Stufe 1 wurden, wenn möglich, Schwellenwerte für Fehlerquoten von 0%, 1%, 5% und 10% ermittelt. Dies entspricht eine Precision von 100%, 99%, 95% und 90%. Hierzu wurden die Konfidenzen der jeweiligen Art aus Stufe 1 absteigend sortiert und durchlaufen, bis die Anzahl falsch-positiver Detektionen die Fehlerquote nicht mehr unterschritt. Des Weiteren wurde die Precision über alle validierten Detektionen einer Art berechnet. Dieses Maß gilt neben dem Recall als eines von zwei

zentralen Gütemaßen von KI-Modellen. Es bildet den Quotienten aus Echt-Positiv-Nachweisen mit allen validierten Nachweisen einer Art. Der Recall wurde im Rahmen der Pilotstudie nicht berechnet, da dieser auch eine Angabe von Falsch-Negativ-Nachweisen benötigt, die im Rahmen der Validierung nicht erhoben wurden.

Die resultierenden Artengemeinschaften auf Grundlage der Validierungen der Stufe 1 wurden mittels Nichtmetrischer Multidimensionaler Skalierung (NMDS) multivariat ausgewertet. Hierzu wurden vier Teildatensätze gebildet, um den Effekt von Falsch-Positiv Nachweisen auf die ökologische Aussage (Unterschiede von Artgemeinschaften) hin zu vergleichen. Der erste Datensatz sollte nur korrekt bestimmte Detektionen beinhalten. Der zweite Datensatz sollte im Gegensatz nur falsch bestimmte Detektionen umfassen. Der dritte Datensatz sollte schließlich sämtliche validierten Daten beinhalten und der vierte Datensatz nur Daten mit einer hohen Konfidenz von pauschal 0,8, ohne Berücksichtigung des Validierungsergebnisses. Der Hintergrund für die Verwendung des vierten Datensatzes war, dass viele Studien pauschale Schwellenwerte nutzen, ohne eine umfängliche Validierung der BirdNET Ergebnisse durchzuführen (vgl. z.B. Sethi et al. 2021). Die NMDS wurde mit dem *vegan* R Paket gerechnet.

Zusätzlich wurde eine ABC-Klassifizierung für jede Art vorgenommen. Dies soll im weiteren Projektverlauf helfen zu ermitteln, für welche Arten eine (teil-)automatische Untersuchung möglich ist. Bei Arten der Klasse C ist dies nicht der Fall. Dies sind entweder von BirdNET detektierte Arten, die nicht bestätigt werden konnten, oder Arten, bei denen sich die Konfidenzen bestätigter Nachweise nicht von denen widerlegter Nachweise unterscheiden oder unter diesen lagen. Bei Arten der Klasse B lagen die Konfidenzen bestätigter Nachweise zwar im Durchschnitt über jenen widerlegter Nachweise, jedoch kam es entweder zu einer hohen Variation, oder die Stichprobengröße war zu gering um eine klare Trennung zu erreichen. Für Arten der Klasse A war eine Trennung der Konfidenzbereiche zwischen bestätigten Detektionen und widerlegten Detektionen deutlich, sodass eine Schwellenwertfindung mit einer gewissen Sicherheit stattfinden konnte. Bei Arten, die keine Falsch-Negativ-Nachweise in der Stufe 1 aufwiesen, wurde je nach Stichprobenzahl und Verteilung der Konfidenzwerte eine Zuordnung in Kategorie A oder B vorgenommen.

Letztlich wurden auch Vergleiche von Validierungen aus Stufe 1 und 2 durchgeführt. Mittels einfacher Netzwerkanalysen im *igraph* R Paket wurden Artnetzwerke von korrekt bestimmten Artnachweisen zu den jeweiligen Begleitarten gebildet. Dies wurde ebenso für falsch bestimmte Artnachweise durchgeführt. Mittels dieser (Fehl)-Bestimmungs-Netzwerke sollen Zusammenhänge aufgedeckt werden, zwischen welchen Arten besonders häufig Fehlbestimmungen auftreten.

# Ergebnisse

## Gerätelaufzeiten

Im aktiven Betrieb lieferten 14 Stationen seit dem 15.4.2023 kontinuierlich Daten. Mangels kontinuierlicher Stromversorgung konnte die Station *Zierenberg* erst ab dem 20.5.2023 Daten liefern, nachdem ein Stromzugang verlegt wurde. Im Zeitraum vom 20.5.2023 bis 26.9.2023 liefen 15 Stationen parallel im Projekt. Am 26.9.2023 musste die Station in *Bad Arolsen* den Betrieb einstellen, weil die dortige Luftmessnetzstation DEHE046 ersatzlos deinstalliert wurde. Somit liefen vom 26.9.2023 bis zum offiziellen Ende der Pilotphase am 30.10.2023 14 Stationen im Dauerbetrieb.

Durch die regelmäßige Internetverbindung mit den ecoSNOOPER konnten auftretende Fehlfunktionen z.T. bereits im Vorfeld erkannt und zeitnah behoben werden. Kleinere Interventionen konnten dabei online, d.h. ohne Anfahrt an die betreffende Station, über die MDM Software geregelt werden. Neben der Erstinstallation erfolgten direkte Anfahrten innerhalb des Projektzeitraums zur Problembeseitigung nur in zwei Fällen.

An der Station *Kellerwald* wurde am 20.5.2023 eine Anfahrt durchgeführt, bei der die interne Laderegulierung neu eingestellt werden musste, nachdem das Gerät einige Tage zuvor durch einen erheblich gestiegenen Energieverbrauch auffiel. Hierdurch reichte die voreingestellte Laderegulierung nicht mehr aus, um das Gerät dauerhaft mit Energie zu versorgen. Als Grund hierfür wird eine länger andauernde Schlechtwetterperiode vermutet, wodurch beim Aufbau der mobilen Datenverbindung mutmaßlich mehr Energie verbraucht wurde. Durch die rechtzeitige Vorwarnung konnte eine Servicefahrt geplant und somit die Ausfallzeit auf wenige Stunden reduziert werden.

Im zweiten Fall wurde an der Station *Kleiner Feldberg* ein plötzlicher Geräteausfall festgestellt, der eine zeitnahe Servicefahrt nötig machte. Das Gerät wurde am 12.10.2023 durch ein Ersatzgerät ausgetauscht. Der Grund für das Herunterfahren von HE16 ist unklar.

Eine weitere systematische Intervention wurde vorsorglich für den Bundesweiten Warntag vorbereitet, der im Projektzeitraum am 14.9.2023 um 11:00 Uhr stattfand. Hierzu wurden im Vorfeld die Geräte so umprogrammiert, sodass der Aufnahmestart um einige Minuten versetzt wurde. Eine mögliche Störung der Gerätefunktion durch den Cell Broadcast Alarm konnte so vermieden werden. Nach Abschluss des Warntages wurden die Geräte mit der nächsten Remote Config Anfrage wieder auf ihren ursprünglichen Aufnahmeintervall zurückgesetzt.

## Artzahlen & Validierungen

Insgesamt lieferten alle Geräte im Projektzeitraum 71632 Einzelaufnahmen, die einer Gesamtaufnahmedauer von 11939 Stunden entsprechen. In diesen Aufnahmen wurden 252 Vogelarten durch BirdNET detektiert. Nach Filterung der initialen Ergebnisse auf Detektionen mit Tagesmaximalkonfidenzen je Station verblieben 96068 Detektionen im Datensatz für die weitere Auswertung. Von diesen Detektionen wurden im Zeitraum vom 15.4.2023 bis 30.7.2023 15033 Detektionen zur Validierung ausgewählt (s.o.), wodurch 118 Arten bestätigt werden konnten. Aus Offenlandlebensräumen wurden insgesamt 101 Arten, mit im Mittel 49 Arten je Standort positiv validiert. Die Vorkommen erstreckten sich von 21 Arten am Standort *Wasserkuppe* bis auf 63 Arten am Standort *Spessart*. An Waldstandorten konnten insgesamt 90 Arten positiv validiert werden. Hier waren im Mittel 55 Arten je Standort anzutreffen, mit einer minimalen Anzahl an der Station *Kleiner Feldberg* mit 44 Arten und 63 Arten an der Station *Fürth/Odenwald*. In urbanen Lebensräumen wurde mit insgesamt 78 Arten die geringste Artenzahl in den drei Habitattypen nachgewiesen. Die mittlere Artenzahl lag bei 48 je Standort jedoch vergleichbar hoch, wie im Offenland. Mit 37 Arten wurden an der Station *Marburg* die wenigsten Arten im urbanen Habitat bestätigt. An der Station *Wiesbaden Süd* konnten mit 62 bestätigten Arten jedoch eine vergleichbar hohe Artenzahl nachgewiesen werden, wie an den besten Standorten im Wald und Offenland (jeweils 63 Arten). Tabelle S1 gibt einen umfassenden Überblick für jede Station.

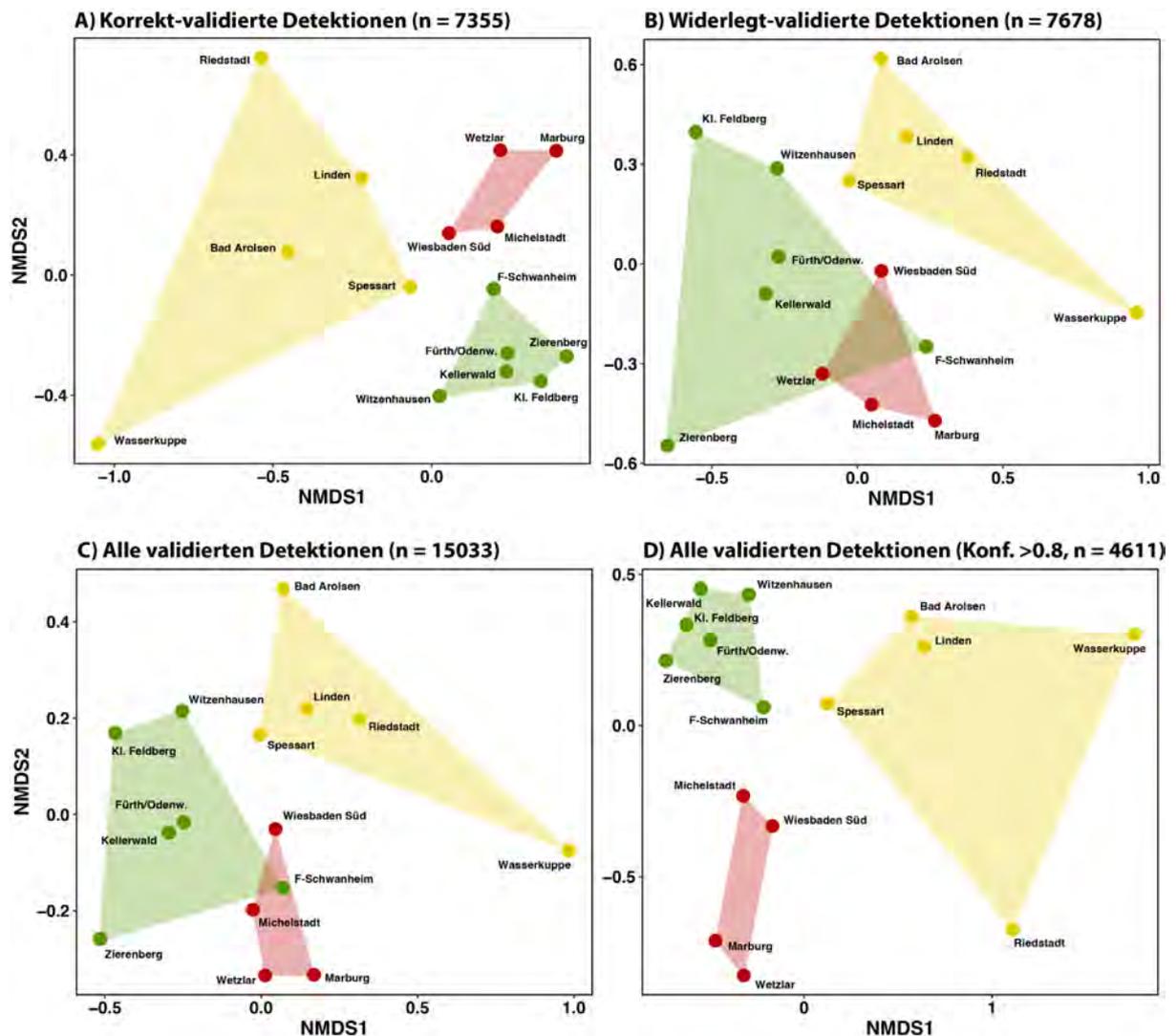


Abb. 2: Ordinationsdiagramm (Nichtmetrische Multidimensionale Skalierung, NMDS) der detektierten und validierten Artengemeinschaften aller Stationen, gruppiert nach Offenland (gelb), urban (rot) und Waldstandorten (grün). Vergleichend dargestellt sind verschiedene Datensätze auf die die NMDS angewendet wurde: A) nur als korrekt-validierte Detektionen, B) nur als widerlegt-validierte Detektionen, C) sämtliche validierte Detektionen, und D) alle Detektionen mit einer Konfidenz über 0.8 unabhängig des Validierungsergebnisses.

Fig. 2: Ordination diagram (Non-metric Multidimensional Scaling, NMDS) of the detected and validated species communities of all stations, grouped by open land (yellow), urban (red), and forest locations (green). Comparatively presented are various datasets on which NMDS was applied: A) only as correctly validated detections, B) only as refuted validated detections, C) all validated detections, and D) all detections with a confidence value above 0.8 regardless of the validation result.

Im multivariaten Vergleich der detektierten Artengemeinschaften mittels NMDS fällt die strikte Trennung nach Habitattyp auf (Abb. 2). Diese Trennung ist jedoch bei Datensätzen, die einen hohen Anteil falsch-positiver Nachweise beinhalten (Abb 2B und 2C) weniger deutlich als bei Datensätzen mit einem hohen Anteil echt-positiver Nachweise (Abb. 2A und 2D). Auch weisen bei Datensätzen mit hohen echt-positiv Nachweisen Offenlandstandorte die größte Variation

auf, wohingegen Stationen im Wald und in urban geprägten Gebieten innerhalb ihres Habitattyps sehr ähnlich sind. Wird ein hoher Anteil falsch-positiver Nachweise berücksichtigt, vergrößert sich der umspannende Ordinationsraum vor allem an Waldstandorten, was auf einen sehr diversen Anteil falsch-positiv bestimmter Arten hindeutet und Interpretationen zur Struktur der Artengemeinschaften verfälscht. Nicht zuletzt auch deshalb, weil sich urbane und Waldstandorte nun zum Teil überlappen (Abb. 2B & C). Interessanterweise sind die Veränderungen im umspannenden Ordinationsraum von Offenlandstandorten nahezu identisch bei Berücksichtigung falsch-positiver Nachweise (Abb. 2B & C). Zwar könnte in diesem Fall über alle Teildatensätze für die Offenlandstandorte der selbe Rückschluss gezogen werden, jedoch zeigt die unterschiedliche Lage der einzelnen Offenlandstandorte im Ordinationsraum, dass die Einbindung von Falsch-Positiv Nachweisen eine fälschliche Stabilität suggerieren, wohingegen die echt-positiv Nachweise recht sensibel darauf reagieren welche Konfidenzen für die Berücksichtigung in die Aufnahme der Analyse gewählt werden. Auch wenn eine pauschal angesetzte hohe Konfidenz aller Detektionen im Vergleich zur aufwändigen Überprüfung von Detektionen vor einer Analyse ein sehr ähnliches Bild liefert (Abb. 2A & D). So scheint es für einen Vergleich einzelner Stationen u.U. eine wichtigere Rolle zu spielen detaillierte Validierungsergebnisse zur Hand zu haben.

Auf Basis der ABC Bewertung der Einzelarten auf Grundlage der Validierungen (s.o.) konnten 63 Arten der Kategorie "A" zugeordnet werden. Weitere 24 Arten erhielten eine "B" Bewertung. Diese 87 A und B Arten eignen sich prinzipiell für zukünftige (teil-)automatische Analysen mit A Arten, die eine höhere Trennschärfe zwischen echt-positiven Detektionen und falsch-positiven Detektionen aufwiesen, und B Arten, die durch eine Vergrößerung der Stichprobe zu einer besseren Trennschärfe führen könnten, oder mit konservativen Schwellenwerten bereits jetzt eine (teil-)automatische Erfassung erlauben. Unter C Arten fallen neben sämtlichen nicht bestätigten Arten auch Arten, die eine zu geringe Stichprobe korrekt bestimmter Nachweise aufwiesen oder deren Konfidenz sich nicht von den Konfidenten widerlegter Detektionen unterschieden. Tabelle S2 gibt genaue Informationen je Art zur Anzahl der durchgeführten Validierungen der Stufe 1.

### Ausgewählte Artbeispiele

Die fünf am häufigsten detektierten Arten waren Blaumeise, Rotkehlchen, Kernbeisser, Kohlmeise und Amsel - allesamt Arten mit Precision Werten zwischen 90% und 100% (Tab S2). Neben diesen generellen Ergebnissen ist die Betrachtung einzelner Arten von besonderer Bedeutung - nicht zuletzt, um auch die Stärken und Schwächen der aktuellen Auswertungstechniken im Detail zu ergründen. Vor allem die Wahl der Schwellenwerte kann

zu sehr unterschiedlichen Eindrücken der räumlichen und zeitlichen Verteilung der Nachweise führen.

Viele Langstreckenzieher haben nur ein sehr kurzes Zeitfenster in dem sie gesangsaktiv sind. Für viele Grasmückenarten liegen die Wertungsgrenzen daher nur zwischen Anfang Mai und Anfang Juni (Südbeck et al. 2005). Am Beispiel der Dorngrasmücke ist dies sehr gut zu erkennen (Abb. 3). Der Nachweisbereich der Art erstreckte sich von Ende April bis Anfang Juli. Ein deutlicher Schwerpunkt der räumlichen Nachweisintensität lag dabei im Mai, und hier vor allem in der ersten Maiwoche. Aus den Validierungen konnte die Art an neun Stationen bestätigt werden, wobei an sechs Standorten mehr als ein Nachweis erfolgte. Entsprechend der Validierungen ergab sich für die Art ein 100% Precision-Schwellenwert bei einer Konfidenz von 0,38 und einer gesamten Precision von 86%. Bei Anwendung dieses Schwellenwerts zeigen sich insbesondere am Freitag und Samstag der ersten Maiwoche eine hohe Nachweiszahl an jeweils sieben Stationen. Dies könnte der Hauptdurchzugszeit der Art in diesem Jahr entsprechen. Im Vergleich zu einem deutlich konservativen Schwellenwert von 0,8 (Abb. 3 unten) bestätigt sich der gesangliche Schwerpunkt der Art im Mai. Hier wird die Art regelmäßig an drei bis vier Stationen pro Tag nachgewiesen. Tatsächlich konnte die Art an vier Stationen mit 11 bis 14 positiven Validierungen bestätigt werden. Aufgrund der Beschränkung von Validierungen auf Tagesmaximalkonfidenzen sind diese Werte mit Erfassungstagen gleichzusetzen, an denen die Art nachgewiesen werden konnte. Die waren mit den Stationen *Kellerwald*, *Witzenhausen*, *Bad Arolsen* und *Spessart* je zwei Wald- und Offenlandstandorte. Mit bestätigten Nachweisen an neun Tagen konnte der urbane Standort in *Wiesbaden-Süd* auch die Dorngrasmücke bestätigen. Jedoch zeigen die beiden Schwellenwerte auch deutliche Unterschiede auf. So sind beim Schwellenwert von 0.38 die letzten beiden Aprilwochen an Einzeltagen mit Nachweisen an einem Standort besetzt, zudem kam es zu sehr späten Detektionen im Oktober. Während die April-Nachweise durchaus frühe Nachweise an einem Standort darstellen könnten, sind die Oktober-Nachweise sicherlich als falsch-positiv zu werten. Ein hoher Schwellenwert bei 0,8, wie er etwa bei Sethi et al. (2021) pauschal Anwendung fand, führt jedoch zu großen Erfassungslücken im Juni. Es ist wahrscheinlich, dass der Gesang der Dorngrasmücke mit einer höheren Konfidenz erkannt wird, als deren Rufe oder Rufreihen. Daher kann eine hohe Filterung vor allem die gesangsintensive Periode im Mai hervorheben, wenn Territorien etabliert und Partner beworben werden. Obwohl beide Schwellenwerte auf Grundlage der durchgeführten Validierungen eine Precision von 100% liefern, sind die artökologischen und jahresphänologischen Interpretationen durchaus unterschiedlich.

Dies zeigt sich noch deutlicher bei häufigen Arten, die sich vergleichsweise leicht bestimmen lassen, wie bei der Blaumeise (Abb. 4). Die Blaumeise gilt in der Studie als häufigste und verbreitetste Vogelart. Sie wurde im Projektzeitraum an über 2300 Studententagen von BirdNET detektiert. Von 353 durchgeführten Validierungen wurde die Blaumeise nur acht Mal widerlegt, woraus sich eine Precision von 98% ergibt. Ein 100% Precision Schwellenwert lag bei der Blaumeise bei einer Konfidenz von 0,5. Legt man diese zugrunde, ergibt sich eine konstant hohe Nachweisdichte zwischen 10 und 13 Stationen pro Tag. Aus den tatsächlich validierten Positiv-Nachweisen wurde die Blaumeise lediglich an der *Wasserkuppe* und in *Riedstadt* jeweils nur einmalig nachgewiesen. Die übrigen Standorte konnten 12 bis 55 bestätigte Nachweistage aufweisen.

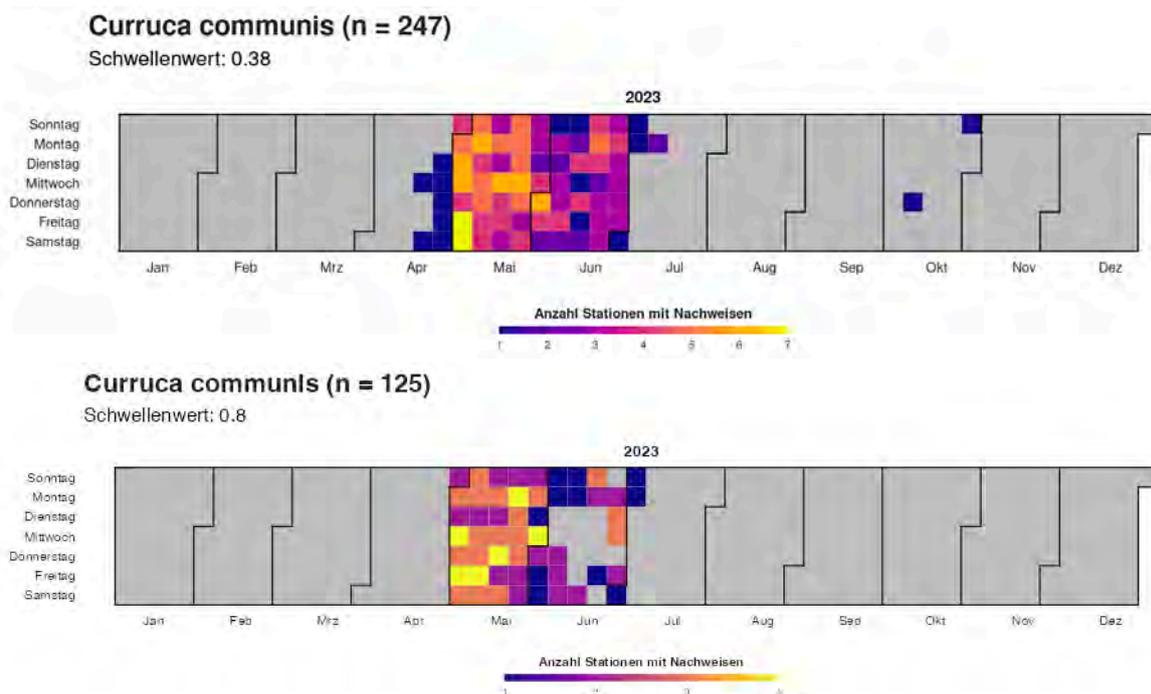


Abb. 3: Aktivitätszeiten der Dorngrasmücke zwischen 15.4.2023 und 30.10.2023, festgestellt anhand des minimalen Schwellenwertes mit 0% falsch-positiver Detektionen (oben, auf Grundlage von 81 validierten Detektionen) und bei pauschaler Konfidenz von 0,8 (unten).

Fig. 3: Activity plot of the Common Whitethroat between April 15th and October 30th, determined based on the minimum threshold value with 0% false positives (top, based on 81 validated detections) and at a fixed confidence of 0.8 (bottom).

Wie schon bei der Dorngrasmücke führt eine Höherlegung des Schwellenwertes jedoch zu anderen Eindrücken. Ein Schwellenwert bei 0,8 entspricht etwa dem 75%-Perzentil aller Echt-Positiv-Nachweise der Blaumeise (also 75% der Positiv-Nachweise hatten einen Schwellenwert von 0,8 oder höher). Bei Anlegung dieses Schwellenwertes wird eine rapide Abnahme der Nachweisdichte ab Mitte Mai deutlich, welche einhergeht mit den in Südbeck et al. 2005 festgelegten Wertungsgrenzen. Da die Pilotstudie erst zum 15.4.2023 startete, lag

der Beginn bereits mitten in der Hauptbalzzeit der Blaumeise (von Februar bis April). Bei einem Schwellenwert von 0,8 wird die Hauptbalzzeit von der übrigen Aktivität und Anwesenheit der Art deutlich hervorgehoben.

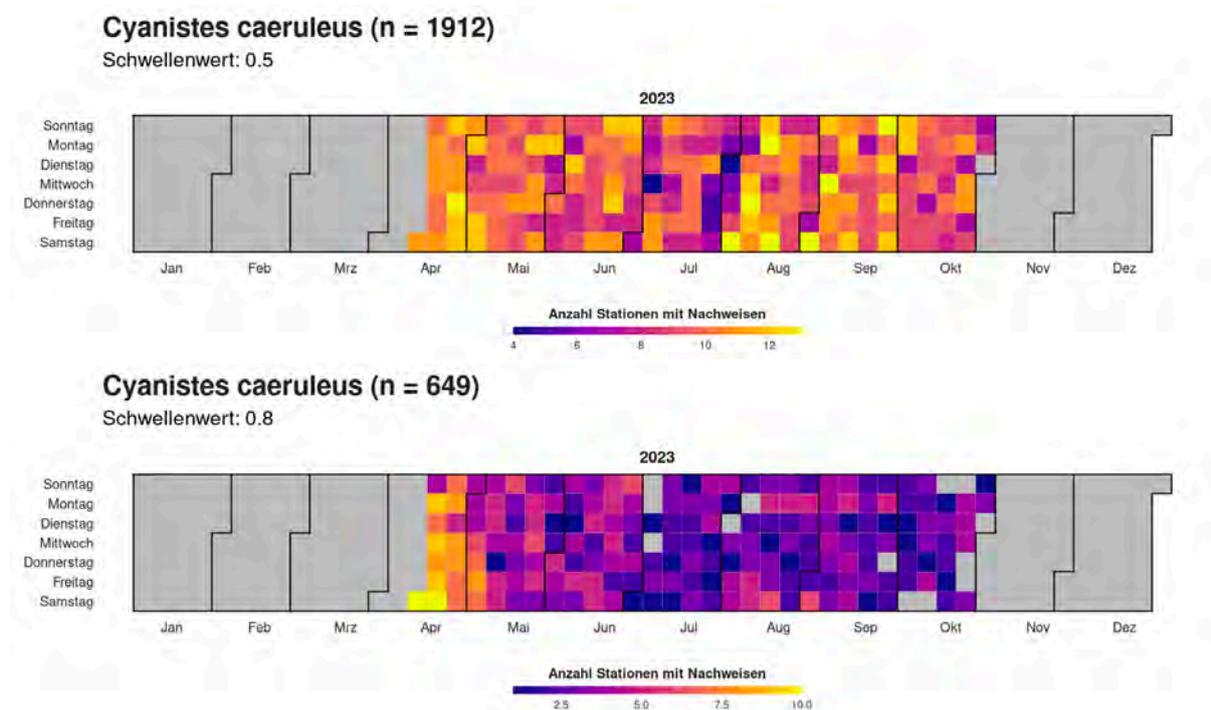


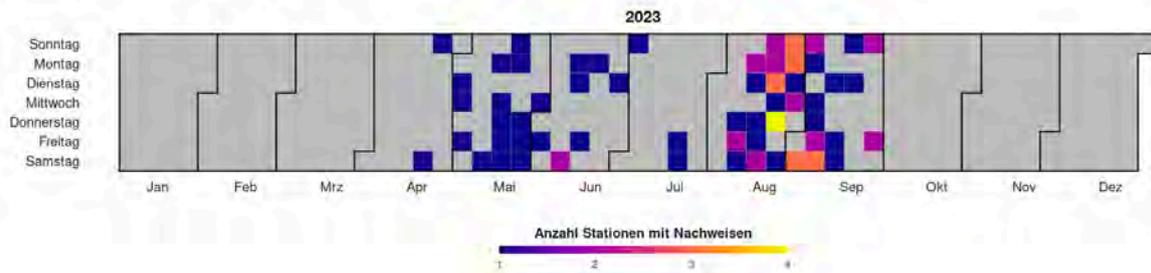
Abb. 4: Beispiel für Unterschiede in der raum-zeitlichen Verteilung detektierter Blaumeisen bei Verwendung unterschiedlicher Schwellenwerte. Der Schwellenwert von 0,5 (oben) entspricht dem niedrigsten Wert, bei dem noch eine Precision von 100% auftrat (auf Grundlage von 353 Validierungen). Der Schwellenwert von 0,8 (unten) entspricht dem 75%-Quartil der als richtig validierten BirdNET Detektionen.

Fig. 4: Example of the differences in the spatial-temporal distribution of detected Blue Tits when using different threshold values. The threshold value of 0.5 (top) corresponds to the lowest value at which a precision of 100% still occurred (based on 353 validations). The threshold value of 0.8 (bottom) corresponds to the 75th percentile of the BirdNET detections validated as correct.

Daher ist es, unabhängig von Validierungen und statistischen Kenngrößen, wichtig, verschiedene Schwellenwerte anzusetzen, um den (Art-)biologischen Erkenntnisgewinn zu maximieren. Da die aktuell verfügbaren Versionen von BirdNET nicht zwischen verschiedenen Typen von Lautäußerung einer Art unterscheiden, ist eine solche Vorgehensweise aktuell die zielführendste wenn man über die alleinige Anwesenheit der Art hinaus biologisch relevante Erkenntnisse aus dem Akustikmonitoring ziehen möchte (mehr auch in den nachfolgenden Abschnitten).

### **Anthus trivialis (n = 83)**

Schwellenwert: 0.875



### **Anthus trivialis (n = 349)**

Schwellenwert: 0.65

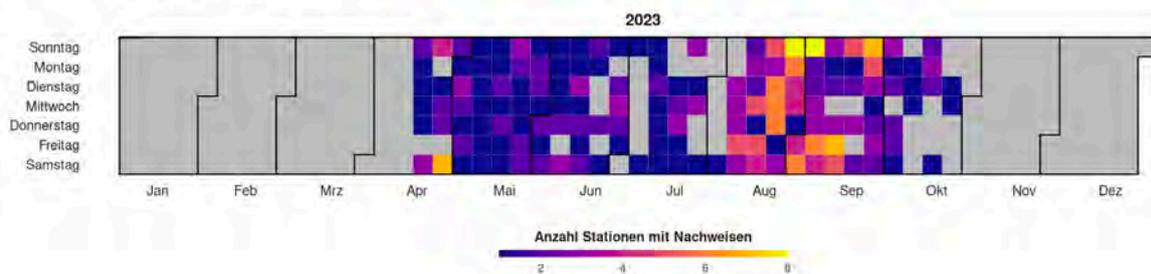


Abb. 5: Beispiel für Unterschiede in der raum-zeitlichen Verteilung detektierte Baumpieper bei Verwendung unterschiedlicher Schwellenwerte. Der Schwellenwert von 0,875 (oben) entspricht dem niedrigsten Wert, bei dem noch eine Falsch-Negativ-Rate von 0% auftrat (auf Grundlage von 202 Validierungen). Der Schwellenwert von 0,65 (unten) entspricht dem einer Precision von 50% Validierter BirdNET Detektionen.

Fig. 5: Example of the differences in the spatial-temporal distribution of detected Tree Pipits when using different threshold values. The threshold value of 0.875 (top) corresponds to the lowest value at which a precision of 100% still occurred (based on 202 validations). The threshold value of 0.65 (bottom) corresponds to a precision of 50% for validated BirdNET detections.

Die differenzierte Einordnung ökoakustischer Ergebnisse ist nicht nur innerhalb der Brutzeit wichtig, sondern auch bei Arten, die sowohl als Brut- als auch als Zugvogel in einem Monitoring auftauchen können. So war in der Pilotstudie vor allem der Herbstzug des Baumpiepers markant (Abb. 5). Im Gegensatz zu den anderen Arten hatte der Baumpieper eine geringe Precision von lediglich 21% bei 202 Validierungen. Aus den Validierungsdaten (Zeitraum 15.4.2023 bis 31.7.2023) konnte die Art an neun Stationen nachgewiesen werden. Mehr als die Hälfte ( $n = 5$ ) der Stationen mit bestätigten Nachweisen hatten jedoch nur Einzelnachweise. Im Gegensatz dazu konnten am Standort *Witzenhausen* 28 Nachweistage aus den Validierungen bestätigt werden. Aufgrund der niedrigen Precision konnte ein 100% Precision-Schwellenwert erst ab einer Konfidenz von 0.875 angesetzt werden. Doch bereits bei diesem hohen Schwellenwert können sowohl der Brutzeitraum von Mai bis in den Juni hinein, vor allem jedoch der Durchzugsschwerpunkt Mitte August bis Mitte September abgegrenzt werden (Abb. 5). Während des Durchzuges konnten in der Spitze bis zu vier Standorte am Tag Nachweise der Art über dem Schwellenwert erbringen. Legt man den Schwellenwert niedriger verliert man in so einem Fall schnell an Precision. Bei einem Schwellenwert von 0,65 lag diese noch bei 50% der validierten Daten. Jedoch scheint sich

dies - zumindest optisch - zugunsten einer besseren Darstellung der Zugperioden zu eignen. Auch wenn mit 202 Validierungen die Überprüfungsleistung weit über dem liegt, was in aktuellen Studien Anwendung findet (z.B. Sethi et al. 2021) könnte sich bei einer Einzelartbetrachtung weiterer Überprüfungsanwendung lohnen, um die (art-)ökologische Aussagekraft weiter zu verbessern. Auch wären hier nach Ruf und Gesang getrennte Classifier für den Baumpieper wünschenswert oder aber zunächst eine saisonale Validierung und Anpassung der Schwellenwerte.

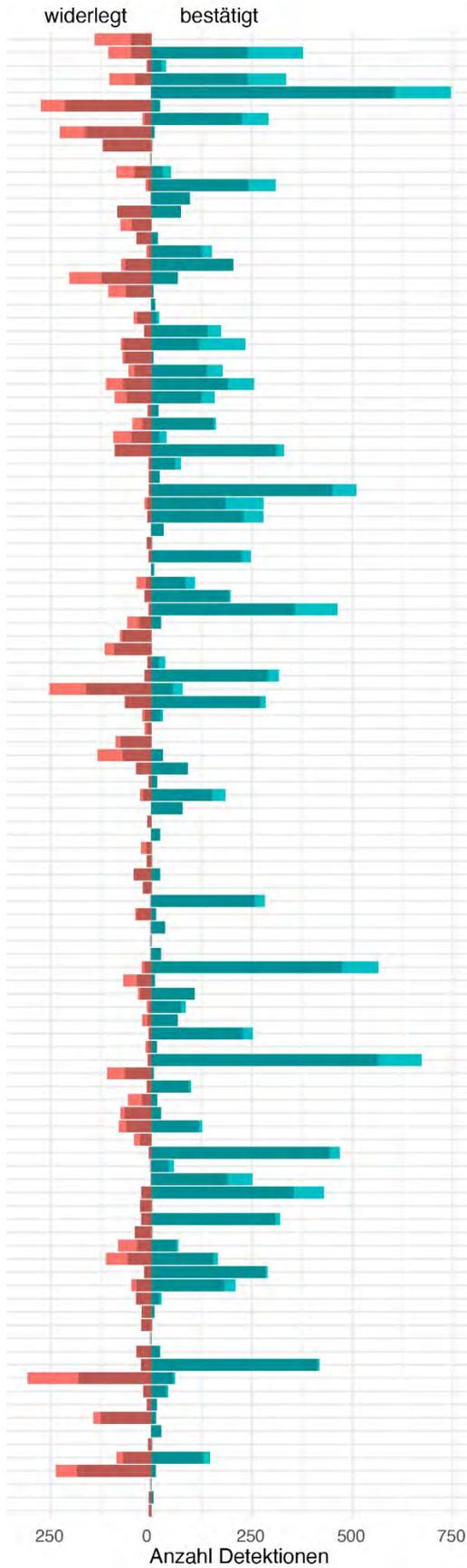
## Begleitarten

Neben der Validierung zur Überprüfung von Detektionen wurde in einem weiteren Schritt kontrolliert, welche sonstigen Detektionen in einem Zeitintervall von neun Sekunden von BirdNET getätigt wurden (vgl. Methoden). Diese detektierten "Begleitarten" waren mit 230 Arten bei falsch-positiven Detektionen wesentlich zahlreicher als bei echt-positiven Detektionen, mit 108 Begleitarten. Hinzu kommen weitere Arten, bei denen es ausschließlich weitere arteigene Detektionen um eine validierte Detektion gab. Diese lagen mit 13 Arten bei Falsch-Positiv-Nachweisen nur geringfügig höher als die 10 Arten bei Echt-Positiv-Nachweisen. Allerdings unterschieden sich falsch-positive Detektionen nicht in der Anzahl der Begleitarten (Mittelwert: 9; 95%KI: 1-22) von echt-positiven Nachweisen (Mittelwert: 11; 95%KI: 1-36). Ein allgemeiner Vergleich von Artnetzen gestaltet sich in der aktuellen Projektphase als schwierig, da durch die unterschiedlichen Anteile zwischen echt-positiven und falsch-positiven Nachweisen absolute Vergleiche, etwa zur Anzahl der Detektionen oder Begleitarten, nur begrenzt möglich sind. Hinzu kommt, dass die Validierung dieser begleitenden Detektionen noch nicht abgeschlossen ist. Eine dezidierte Unterscheidung zwischen echt-positiven und falsch-positiven Begleitarten kann daher aktuell nicht erfolgen. Dennoch soll hier auf eine deskriptive Beschreibung der bisherigen Ergebnisse nicht verzichtet werden.

Arteigene Detektionen um einen bereits validierten Artnachweis der Stufe 1 herum machen beim überwiegenden Teil der untersuchten Arten den Großteil der Begleitarten-Detektionen aus (Abb. 6). Bedenkt man, dass der arteigene Gesang oftmals über die von BirdNET genutzten drei Sekunden langen Betrachtungsfenster hinausgeht, scheint es plausibel, dass dieselbe Art auch in den unmittelbar benachbarten Zeitfenstern detektiert wird. Weitere Vergleiche der Detektionszahlen sind aufgrund der unterschiedlichen Anzahl an Validierungsergebnissen aktuell jedoch nicht möglich. Eine Betrachtung der Anzahl an Begleitarten ist jedoch weniger stark von dieser Einschränkung betroffen (Abb. 6). Abbildung 7 zeigt das Netzwerk aus Begleitarten in kreisförmiger Anordnung für falsch-positive (Abb. 7A) und echt-positive Detektionen (Abb. 7B). Hieraus ergibt sich ein dynamisches Bild. So gibt es

Arten, wie die Amsel, die bei echt-positiven Nachweisen eine höhere Zahl detektierter Begleitarten aufweist als bei falsch-positiven Nachweisen (Abb. 6). Dies könnte ein Hinweis darauf sein, dass es bei Anwesenheit der Amsel zu einer höheren Anzahl von Fehlbestimmungen kommt, wohingegen die Anzahl der Arten, welche zu einer Fehlbestimmung der Amsel führen, geringer ist. Betrachtet man darüber hinaus noch all jene nachweislich falsch-bestimmten Arten, bei denen die Amsel als Begleitart detektiert wurde (Abb. S1), scheint der Einfluss der Amsel auf Falsch-Positiv-Raten anderer Arten sehr groß zu sein. Bei anderen Arten, wie dem Baumpieper, ist die Anzahl der Begleitarten - v.a. auch solche, die häufiger gemeinsam vorkommen (Abb. 6) - bei falsch-positiven Detektionen deutlich höher als bei echt-positiven Detektionen. So wurden als Begleitarten von falsch-positiven Baumpieper-Detektionen besonders häufig Feldlerche, Buchfink und Hausrotschwanz detektiert (Abb. S1). Durch die laufende Validierung dieser Begleit-Detektionen soll hierzu mehr Klarheit gewonnen werden, welche Arten oft verwechselt werden und inwieweit sich diese Kenntnisse nutzen lassen um den Anteil falsch-positiver Nachweise im ökoakustischen Monitoring zu reduzieren.

validierte Zielart



- Vanellus vanellus*
- Turdus viscivorus*
- Turdus pilaris*
- Turdus philomelos*
- Turdus merula*
- Turdus iliacus*
- Troglodytes troglodytes*
- Tringa ochropus*
- Tadorna ferruginea*
- Tachybaptus ruficollis*
- Sylvia borin*
- Sylvia atricapilla*
- Sturnus vulgaris*
- Strix aluco*
- Streptopelia decaocto*
- Spizus spizus*
- Sitta europaea*
- Serinus serinus*
- Scolopax rusticola*
- Saxicola rubicola*
- Riparia riparia*
- Remiz pendulinus*
- Regulus regulus*
- Regulus ignicapilla*
- Rallus aquaticus*
- Pyrrhula pyrrhula*
- Prunella modularis*
- Poecile palustris*
- Poecile montanus*
- Picus viridis*
- Picus canus*
- Pica pica*
- Phylloscopus trochilus*
- Phylloscopus sibilatrix*
- Phylloscopus collybita*
- Phoenicurus phoenicurus*
- Phoenicurus ochruros*
- Phasianus colchicus*
- Pernis apivorus*
- Periparus ater*
- Perdix perdix*
- Passer montanus*
- Passer domesticus*
- Parus major*
- Oriolus oriolus*
- Numenius phaeopus*
- Nucifraga caryocatactes*
- Muscicapa striata*
- Motacilla flava*
- Motacilla cinerea*
- Motacilla alba*
- Milvus milvus*
- Milvus migrans*
- Mareca penelope*
- Luscinia megarhynchos*
- Lullula arborea*
- Loxia curvirostra*
- Lophophanes cristatus*
- Linnaria cannabina*
- Larus argentatus*
- Lanius collurio*
- Hippolais polyglotta*
- Glaucidium passerinum*
- Gallinula chloropus*
- Fulica atra*
- Fringilla coelebs*
- Ficedula hypoleuca*
- Falco tinnunculus*
- Falco subbuteo*
- Falco peregrinus*
- Erithacus rubecula*
- Emberiza schoeniclus*
- Emberiza citrinella*
- Dryocopus martius*
- Dendrocoptes medius*
- Dendrocoptes major*
- Delichon urbicum*
- Cyanistes caeruleus*
- Curruca curruca*
- Curruca communis*
- Cuculus canorus*
- Coturnix coturnix*
- Corvus monedula*
- Corvus frugilegus*
- Corvus corone*
- Corvus corax*
- Columba palumbus*
- Coccythraustes coccythraustes*
- Ciconia ciconia*
- Chloris chloris*
- Charadrius dubius*
- Certhia familiaris*
- Certhia brachydactyla*
- Carduelis carduelis*
- Buteo buteo*
- Branta leucopsis*
- Branta canadensis*
- Athene noctua*
- Ardea purpurea*
- Ardea cinerea*
- Apus apus*
- Anthus trivialis*
- Anthus pratensis*
- Anser anser*
- Anas platyrhynchos*
- Alopochen aegyptiaca*
- Aegolius funereus*
- Aegithalos caudatus*
- Actitis hypoleucos*
- Accipiter nisus*
- Accipiter gentilis*
- Acanthis cabaret*

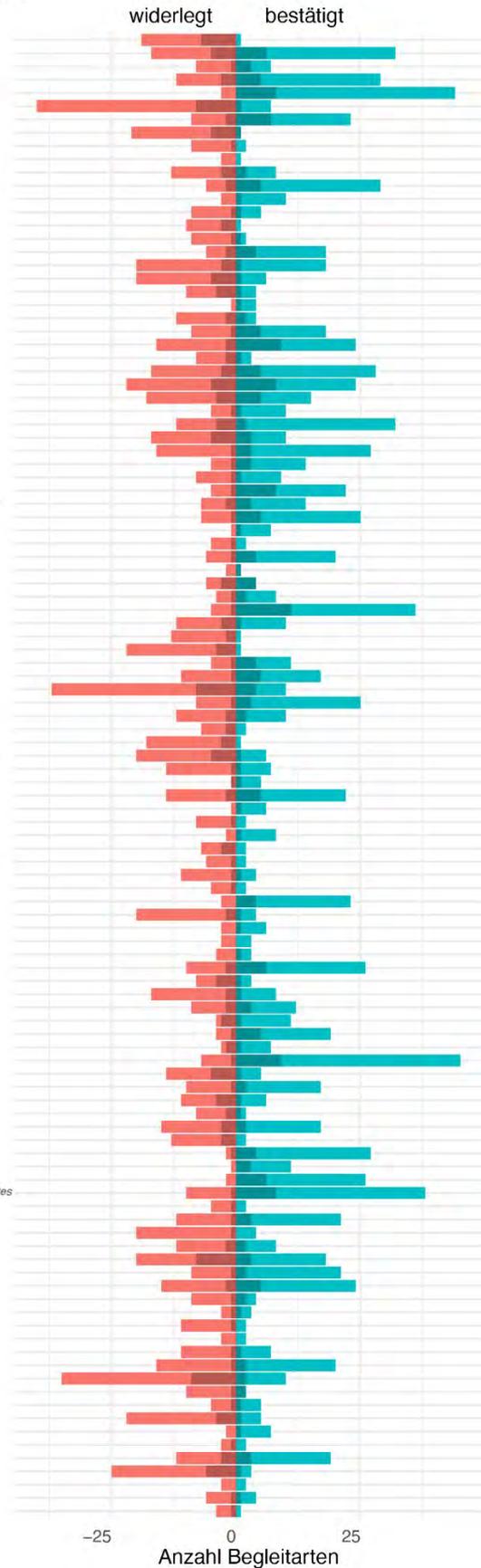


Abb. 6: Charakterisierung der Begleitartenstruktur bei validierten Arten mit Falsch-Positiv- und Falsch-Negativ-Nachweisen ( $n = 112$ ). Links zeigt die Gesamtzahl sonstiger Detektionen um einen validierten Nachweis herum an widerlegten (rot) und bestätigten (blau) Detektionen. Dunkle Schattierungen entsprechen weiteren Detektionen der selben Art (links). Rechts zeigt die Anzahl Begleitarten an widerlegten (rot) und bestätigten (blau) Detektionen. Dunkle Schattierungen entsprechen Arten die mindestens fünf Mal zusammen mit der validierten Art detektiert wurden.

*Fig. 6: Characterization of the accompanying species structure in validated species with false positive and false negative detections ( $n = 112$ ). On the left, the total number of other detections around a validated evidence is shown to refute (red) and confirm (blue) detections. Dark shades correspond to further detections of the same species (left). On the right, the number of accompanying species in refuted (red) and confirmed (blue) detections is shown. Dark shades correspond to species that were detected at least five times together with the validated species.*



Abb. 7: Netzwerk von Begleitarten für A) falsch-positive und B) echt-positive Detektionen. Farbe und Stärke der Verbindung zeigen proportional die Häufigkeit des gemeinsamen Auftretens an. Für eine bessere Übersicht wurden nur Verbindungen zu Begleitarten dargestellt, die mehr als fünf Mal nachgewiesen wurden. Somit stellt die Abbildung 192 von 230 Arten bei falsch-positiven Detektionen bzw. 92 von 108 Arten bei echt-positiven Detektionen dar.

*Fig. 7: Network of accompanying species for A) false-positive and B) true-positive detections. The color and strength of the connection indicate the frequency of co-occurrence proportionally. For better clarity, only connections to accompanying species that were detected more than five times were shown. Thus, the figure represents 192 of 230 species in false-positive detections and 92 of 108 species in true-positive detections.*

## Diskussion

### Tauglichkeit des Systems

Das eingesetzte System um den ecoSNOOPER überzeugte durch eine enorm hohe Zuverlässigkeit, mit einer Ausfallrate von lediglich 0,54% im Projektzeitraum. Der Aufbau eines Frühwarnsystems, welches beispielsweise die Akkuladestände im Blick behält, half bei der Zahl an Geräten den Überblick zu behalten. Dennoch kam es zu kurzfristigen Ausfällen. Zum Teil lag dies an einem erhöhten Stromverbrauch, der durch die analog einzustellende Ladeautomatik nicht mehr ausgeglichen werden konnte. Durch das Unterschreiten der Ladestände und dem negativen Trend konnte eine Intervention jedoch geplant und umgesetzt werden, sodass der zeitliche Ausfall nur wenige Stunden betrug. In einigen Fällen ließ sich durch die Umstellung des Flugmodus temporär Energie sparen, um einem Ausfall entgegenzuwirken. Zwar wird durch eine Maximierung der Zeit im aktiven Flugmodus der Fernzugriff auf das Gerät verhindert, jedoch ergeben sich dennoch kurze Zeitfenster pro Tag in denen Interventionsmaßnahmen möglich sind, wenn ein Fernzugriff auf das Gerät nötig wurde (etwa zum Verschieben und Archivieren der Daten auf die lokale microSD-Karte).

Ein weiterer kritischer Punkt war die Performance der ecoSNOOPER bei hohen Temperaturen im Sommer. So wurde befürchtet, dass in einer Hitzewelle die Wärmeentwicklung innerhalb des Gehäuses einen kritischen Wert überschreiten könnte, der die Geräte in den Not-Aus-Status versetzt, um die Batterie vor kritischer Überhitzung zu schützen. Dieser Fall ist nicht eingetreten. Zwar stiegen die Gerätetemperaturen im Sommer regelmäßig bis auf knapp 50°C - dies hatte aber keine erkennbaren Auswirkungen auf die Gerätefunktion.

Eine letzte offene Frage ist jedoch die Aufnahmeleistung bei starker Kälte. Standorte wie die *Wasserkuppe* oder dem *kleinen Feldberg* sind im Winter regelmäßig gefrierendem Nebel und Eis ausgesetzt. Durch die exponierte Lage, vor allem des Mikrofons, ist zu befürchten, dass sich erhebliche Mengen Eis um bzw. am Mikrofon bilden können und dieses dämpfen, bzw. beschädigen und somit die Aufnahmeleistung erheblich eingeschränkt ist. Wenngleich in

dieser Jahreszeit nicht mit hohen Vogelaktivitäten an den Standorten zu rechnen ist, ist die Frage dennoch wichtig zu beantworten, ob die Geräte eines Dauermonitorings auch den teils harschen Witterungen im Winter widerstehen können. Diese Frage muss im Rahmen des Pilotprojektes noch unbeantwortet bleiben. Es werden jedoch Tests für Verfahren durchgeführt, um einer Eisbildung entgegenzuwirken.

## Schwellenwerte und Validierungen

Die Ermittlung zuverlässiger Schwellenwerte, die für eine automatische Erfassung unabdingbar sind, stellt eine erhebliche Herausforderung dar, da es eine Vielzahl von Stellschrauben gibt, die auf die Erstellung solcher Schwellenwerte wirken. Zunächst mal ist von einer pauschalen Anwendung eines fixen Schwellenwertes über Arten hinweg abzuraten. Beispielsweise nutzten Sethi et al. 2022 im Sound of Norway Projekt eine Konfidenz von 0,8 als pauschale Detektionsschwelle um die Anwesenheit einer Art zu klassifizieren. Anhand der in dieser Studie durchgeführten Validierungen läge die mittlere Detektionsschwelle bei absolutem Ausschluss von Falsch-Positiv-Nachweisen jedoch schon bei 0,65. Betrachtet man die Einzelart, reichen diese Schwellenwerte (bei 0% FP) von 0,31 bis 0,965. Ein pauschaler Schwellenwert von 0,8, wie beispielsweise von Sethi et al. 2022 verwendet, würde insgesamt die Anzahl echter Nachweise erheblich unterschätzen; auf die Einzelart betrachtet jedoch zu erheblichen Unter- oder Überschätzungen führen. So legen die in der vorliegenden Pilotstudie durchgeführten Baumpieper Validierungen (n = 202) einen Schwellenwert von 0,875 nahe, um Falsch-Nachweise zu minimieren. Bei der Blaumeise genügte jedoch schon ein Schwellenwert von 0,5, um bei 353 Validierungen insgesamt sämtliche Falsch-Nachweise (n = 8) zu ignorieren.

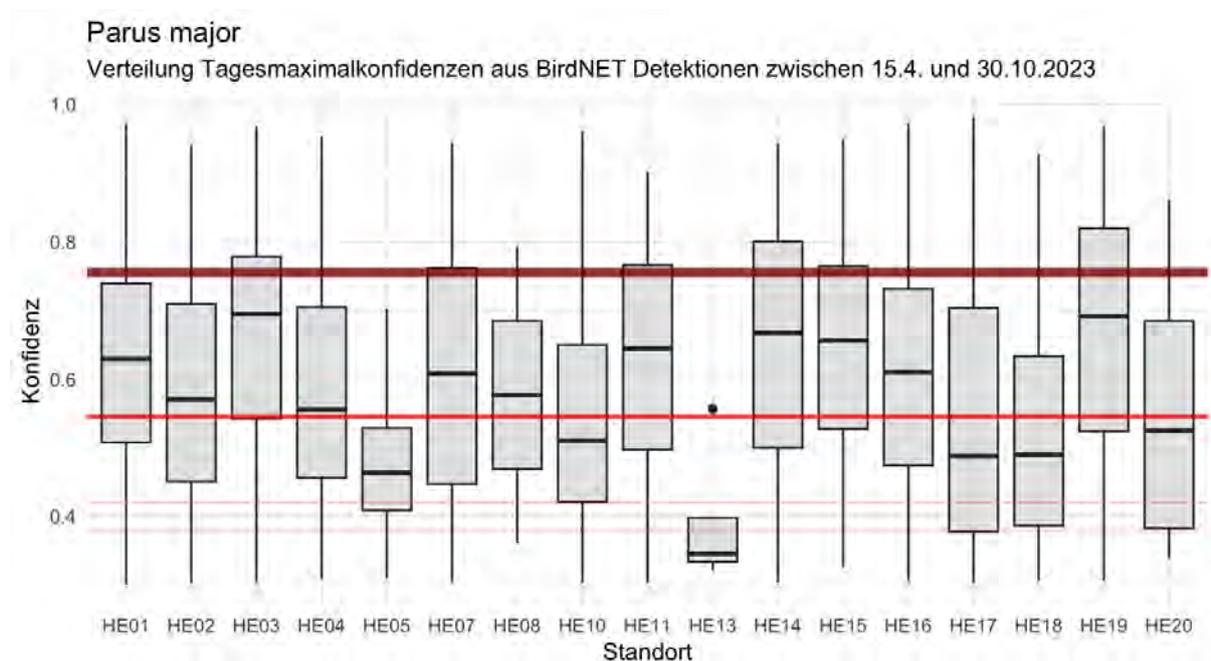
Jedoch bergen die Validierungen an sich auch gewisse Risiken und Unschärfen. Zunächst einmal kann auch bei größter Sorgfalt nicht gewährleistet werden, dass alle Überprüfungen auch tatsächlich der Wahrheit entsprechen. Neben einem hohen Grad an Expertise in der akustischen Bestimmung von Vogellauten spielen auch eine Reihe kognitiver Verzerrungen eine Rolle, die zum einen individuell ausgeprägt sind, zum anderen jedoch auch mit dem Grad der Konzentration zunehmen können. So ist davon auszugehen, dass eine Fehlerrate mit zunehmender Zeit, die jemand mit der Validierung beschäftigt ist, zunimmt. Für eine zuverlässige Datengrundlage wird daher empfohlen, die Validierung nicht von Einzelpersonen durchführen zu lassen, sondern von mehreren Experten und nicht für lange Zeiträume am Stück. Durch die mehrfache unabhängige Validierung unterschiedlicher Experten von einem Teil der Detektionen ließen sich Consensus Bewertungen ableiten, welche sowohl die individuelle Qualität der Experten aufdecken, als auch eine solidere Datenbasis versprechen.

Zudem ist es wichtig, mehr über die Anzahl der Validierungen, die nötig sind, um Schwellenwerte zu identifizieren, zu lernen. Validierungen sind zeit- und damit kostenaufwändig - gerade bei Vögeln, die gemeinhin eine artenreiche Gruppe darstellen. Auf Grundlage der Artenliste Hessens, welche Brut- aber auch Rastvögel beinhaltet, hat BirdNET 252 detektiert. Da diese Studie zum Ziel hatte, Aussagen über sämtliche Vogelarten zu treffen, mussten auch alle 252 von BirdNET detektierten Arten überprüft werden. Es zeigte sich, dass die Anzahl der Validierungen z.T. sehr hoch sein kann und dennoch eine gewisse Unschärfe in der Aussage liefert. Beispielsweise gab es bei Stelzen erhebliche Probleme, Echt-Positiv- und Falsch-Positiv-Nachweise voneinander zu trennen. Bei der Bachstelze wurden 204 Validierungen vorgenommen, von denen 163 als richtig bestätigt werden konnten. Allerdings überschritten sich echt-positive und falsch-positive erheblich in den von BirdNET ausgegebenen Konfidenzen (Abb. 9), sodass ein Schwellenwert mit der Zielsetzung Falsch-Positiv-Nachweise zu minimieren bei 0,925 gesetzt werden musste. Eine hohe Zahl von echt-positiven Detektionen geht damit jedoch verloren, was beispielsweise eine Interpretation (jahres-)zeitlicher Dynamiken erheblich einschränkt.

Ein weiterer wesentlicher Punkt für die Validierung und die damit eingehende Definition von Schwellenwerten ist die Wahl der zu validierenden Detektionen. Viele Studien nehmen dabei eine von zwei Möglichkeiten: Eine gleichmäßige Verteilung von Detektionen entlang eines Konfidenzbereiches aus den BirdNET Einstellungen (also zum Beispiel zwischen 0,3 und 1,0) beziehungsweise daraus abgeleiteter Klassen (z.B. in Schritten von 0,2) oder aber durch eine Vorfilterung eines pauschalen Schwellenwerts (z.B. 0,8; Sethi et al. 2022). Beide Methoden erlauben ein einheitliches Vorgehen und damit eine vermeintliche Vergleichbarkeit der Ergebnisse. Allerdings führt - wie bereits oben gezeigt - eine pauschale Festlegung zu erheblichen Über- oder Unterschätzungen auf Ebene der Einzelart. Hinzu kommt, dass Detektionen einzelner Untersuchungsstandorte spezifisch für die Klanglandschaft des jeweiligen Standortes sind. So kann es sein, dass sich allein durch die Entfernung eines Gesangrevieres die mittlere Konfidenz korrekter Detektionen für eine Art stark unterscheidet. So wären Schwellenwerte je Standort ganz unterschiedlich an die entsprechenden Geräuschkulissen anzupassen. Pauschal gesetzte Konfidenzklassen oder Konfidenz-Untergrenzen ignorieren diesen Umstand was im Zweifel zu einer stark verzerrten Wiedergabe der tatsächlichen Artengemeinschaft führt.

Mit der Zielvorgabe aus diesem Pilotprojekt artspezifische Schwellenwerte zu definieren, erfolgte die Auswahl von Validierungen auf Grundlage von Tagesmaximalkonfidenzen je Art und Standort und mit absteigender Konfidenz der daraus resultierenden Treffer. Dies hatte den Vorteil, dass die zur Verfügung stehenden Ressourcen bestmöglich eingesetzt werden konnten, um einen guten Eindruck von der Schwellenwertverteilung zu erhalten, da viele

suboptimale Detektionen nicht berücksichtigt wurden. Jedoch führt dies zu einer systematischen Überrepräsentation einzelner Geräuschkulissen, in denen eine Zielart möglichst nah zum Akustikrecorder vokal aktiv ist (Abb. 8). Folglich wären Situationen, in denen eine Zielart zwar akustisch nachweisbar ist, sich jedoch weiter entfernt vom Recorder befindet und somit niedrigere Tagesmaximal-Konfidenzen erzeugt, seltener in die Validierung eingegangen. Jedoch könnte im Sinne einer Punkt-Erfassung argumentiert werden, dass weit entfernte Individuen aus der Erfassung auszuschließen sind (was durch den Einsatz von fixen Radien bei der Punkt-(Stopp-)Erfassung gewährleistet wird). Die Aussagekraft nahegelegener Artnachweise zu schärfen, könnte daher dem akustischem Equivalent einer Punkt-(Stopp-)Erfassung zugutekommen. Alternativ könnte in zukünftigen Erfassungen eine absteigende Anordnung der Tagesmaximal-Konfidenzen pro Aufnahmestation erfolgen und so Ressourcen zur Validierung gleichmäßig über alle Stationen aufgeteilt werden.



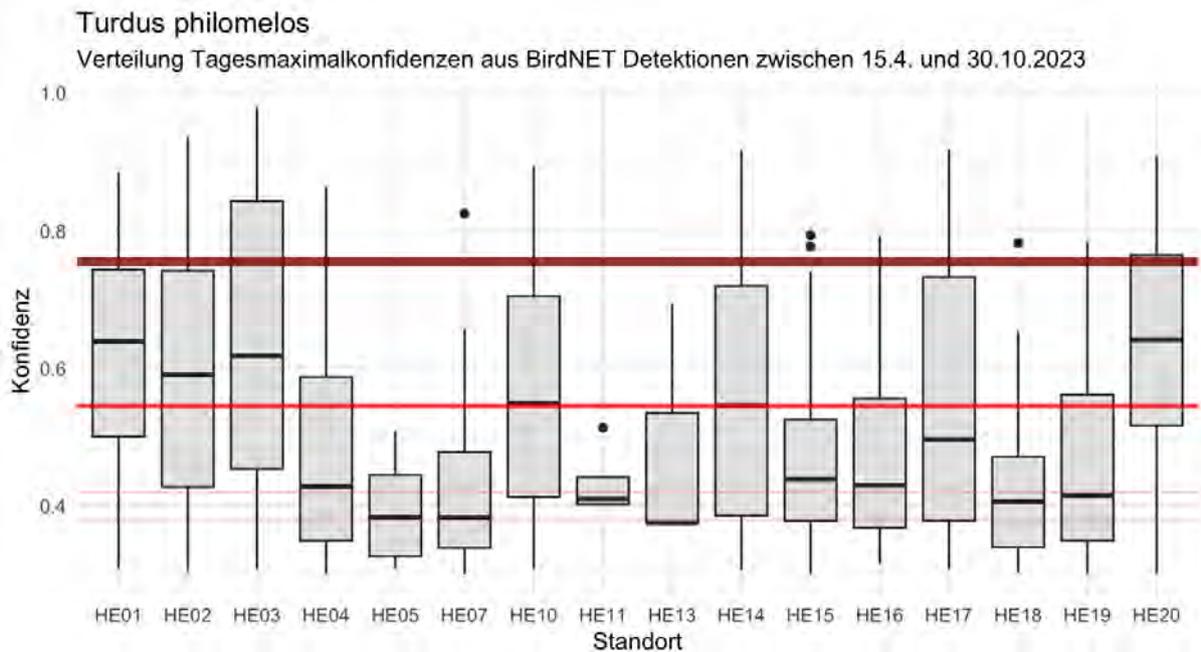


Abb. 8: Exemplarische Darstellung der Verteilung von Tagesmaximalkonfidenzen von Kohlmeise (oben) und Singdrossel (unten) je Aufnahmeort. Horizontale Linien geben die aus den Validierungen ermittelten Schwellenwerte nach der in Tab S2 definierten Akzeptanz von Falsch-Positiv-Nachweisen an. Dunkelrot = 0% Falsch-Positive, Rot (dick) = 1% Falsch-Positive, Rot = 5% bzw 10% Falsch-Positive.

*Fig. 8: Exemplary representation of the distribution of daily maximum confidences of Great Tit (above) and Song Thrush (below) per recording location. Horizontal lines indicate the threshold values determined from the validations according to the acceptance of false positives defined in Table S1. Dark red = 0% false positives, Red (thick) = 1% false positives, Red = 5% or 10% false positives.*

Als letzter Punkt zum Thema Validierungen und Bewertung der Ergebnisse ist die Kenngröße der Precision. Dieser Wert wird - wie bereits oben näher beschrieben - standardmäßig bei der Bewertung der Bestimmungsleistung von akustischer Vogelbestimmungs-Algorithmen herangezogen. Pérez-Granados (2023) gibt einen aktuellen Überblick im Bezug auf BirdNET. Der Autor gibt bei seiner Literaturstudie zu bedenken, dass Precision Werte sich nicht pauschal übertragen, oder gar zwischen Studien vergleichen lassen, da die Berechnung der Precision auf unterschiedlichen Wegen erfolgt. Beispielsweise nutzten Sethi et al. 2021 in der Sound of Norway Studie nur Arten mit mehr als 50 Detektionen mit einer Konfidenz >0,8 für die Berechnung der Precision. Dies ignoriert zum einen Arten mit schlechter Performance und überschätzt gleichzeitig die Precision bei den verbleibenden Arten. In der vorliegenden Pilotstudie wurde die Precision für alle validierten Arten berechnet (entspricht dem Konfidenzbereich 0,3 bis 1,0). Über all diese Arten (n = 15.033 Validierungen) lag die Precision bei 49%. Allerdings ist durch die Wahl der Detektionen mit Blick auf absteigende Tageshöchstkonfidenzen je Art und Standort die Precision ebenso schwer mit anderen Studien zu vergleichen, die ein anderes Auswahlverfahren zur Validierung angewandt haben. Pérez-Granados (2023) diskutiert in diesem Zuge auch die artspezifischen Unterschiede der Precision zwischen Studien. Am Beispiel des Kolkraben variierte die studienspezifische

Precision zwischen 0,29 und 0,94 (Pérez-Granados 2023 und Referenzen ebenda), lag in dieser Pilotstudie dennoch bei 0,89. Neben Unterschieden in der Auswahl der validierten Detektionen spielt hier sicherlich auch das Vorkommen und Auftreten im Detektionsradius des Aufnahmegegeräts eine wichtige Rolle bzw. die lokale Klangkulisse. So lag der Schwellenwert beim Kolkkraben ohne Falsch-Positiv-Nachweise bereits bei niedrigen Konfidenzen von 0,395. Auch sind Aussagen zur Precision bei kleinen Stichproben kritisch zu betrachten. Ein wichtiger Aspekt, der bislang nicht in der internationalen Literatur behandelt worden zu sein scheint, ist, dass sich die Precision nicht pauschal als Aussage über die Nutzbarkeit des Algorithmus zur Artbestimmung eignet. In der Pilotstudie konnte die Bachstelze beispielsweise eine Precision von sehr guten 80% erreichen. Hingegen reichte es bei der Gebirgsstelze nur für 20%. Jedoch wiesen beide Arten eine breite Streuung der Detektionen über den berücksichtigten Konfidenzbereich zwischen 0,3 und 1,0 auf - sowohl für die echt-positiven als auch für die falsch-positiven Nachweise (Abb. 9). So kann zwar, wie bei der Bachstelze, der gesamte Anteil echt-positiver Nachweise hoch sein, jedoch geht dies nicht gleich mit einer guten Trennung von korrekt und falsch detektierten Nachweisen einher. Umgekehrt gilt es, dass sich bei einem niedrigen Anteil echt-positiver Nachweise einer Art, wie bei der Gebirgsstelze, dennoch Erkenntnisse gewinnen lassen (Abb. 9).

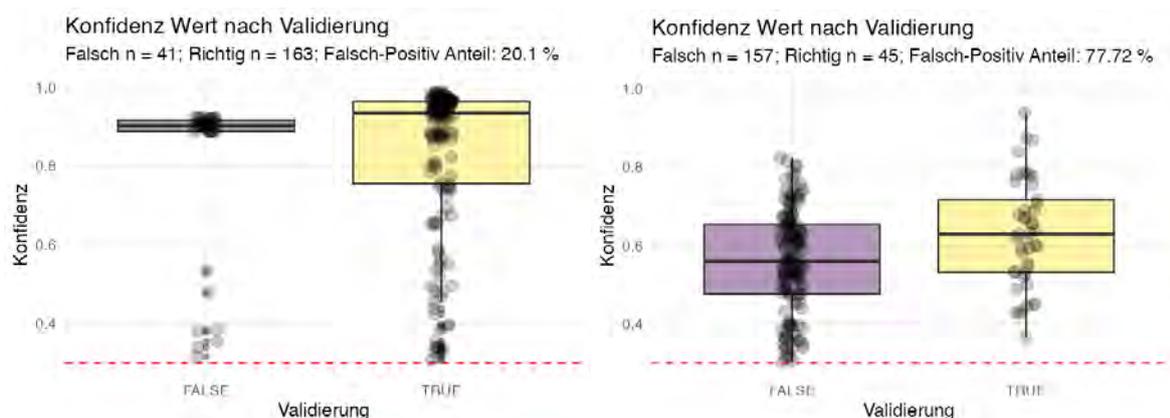


Abb. 9: Verteilung Echt-Positiver- (gelb) und Falsch-Positiver (lila)- Detektionen (Punkte) von Bachstelze (links) und Gebirgsstelze (rechts). Die gestrichelte Linie liegt bei der minimal berücksichtigten Konfidenz von 0,3.

Fig. 9: Distribution of True Positive (yellow) and False Positive (purple) detections (dots) of Pied Wagtail (left) and Grey Wagtail (right). The dashed line is at the minimum considered confidence of 0.3.

Somit ist die Precision als Gütemaß durchaus kritisch zu betrachten, gerade im Hinblick auf Vergleiche zwischen Arten und Studien, aber auch generell als Entscheidungshilfe zur Dateninterpretation und Modellperformance. Tabelle S2 führt Schwellenwerte für jede Art bei einer Precision von 100%, 99%, 95% und 90% auf.

# Ausblick

Die vorliegende Pilotstudie konnte zeigen, dass bei dauerhafter Stromversorgung eine zuverlässige responsive Hardware in der Lage ist ökoakustische Daten mit einer hohen Abdeckung zu erzeugen. Die aktuell gängigen Analysesysteme sind sicher bei weitem noch nicht perfekt und nur mit entsprechender Nachbereitung für weitere Analysen verwendbar. Jedoch zeigt das Feld der künstlichen Intelligenz in rasanter Geschwindigkeit enorme Entwicklungssprünge, sodass man davon ausgehen darf, dass sich im Bereich der automatischen Datenauswertung ökoakustischer Daten noch viel bewegen wird. Auf Grundlage der hier gewonnen Erkenntnisse und der generellen Weiterentwicklung in dem Forschungsfeld soll hier auf zwei Themen ein Ausblick gerichtet werden.

## Langlebigkeit ohne Stromanschluss

Die Verfügbarkeit eines permanenten Stromanschlusses war eine der wesentlichen Voraussetzungen diese Pilotstudie umzusetzen. Jedoch sind viele der gerade ornithologisch spannenden Flächen nicht in unmittelbarer Nähe einer Steckdose zu finden. Eine Vielzahl von Systemen erreicht bereits lange Laufzeiten mit Akkulösungen, jedoch mit der Kehrseite, dass Daten nicht abrufbar und Geräte nicht überprüfbar sind, was deren Nutzen in vielen Bereichen stark einschränkt - gerade wenn ökoakustische Daten rechtssichere Aussagen zum Vorkommen relevanter Arten ermöglichen sollen. Aus diesem Grund testen wir verschiedene Akkulösungen für den ecoSNOOPER, um lange Standzeiten mit geringem Wartungsaufwand vor Ort zu erreichen. Das Basisgerät des ecoSNOOPERs kann mit dem geräteinternen Akku bereits eine Standzeit von ca. 3 Wochen überbrücken, wenn man die energiehungrigen Einstellungen aus dem Dauermonitoring zugrunde legt. (So kann ein ecoSNOOPER bei acht Stunden Aufnahme am Tag, verteilt auf Vormittag und Abend, und ohne Datentransfer durchaus 4-5 Wochen im Gelände aktiv bleiben.) Wir erweiterten die Akkuleistung des ecoSNOOPERs mit dem Einbau eines externen Akkus der in dem selben Gehäuse Platz findet. Nach verschiedenen Durchläufen und Geräteeinstellungen mit verschiedenen Laderegeln schafft die aktuelle Akku-Variante zwei Monate mit demselben Zeitplan wie die strombetriebenen Geräte. Der einzige Unterschied ist, dass Daten nicht übertragen werden, ein Fernzugriff auf das Gerät jedoch bei bestehender Mobilfunkverbindung ermöglicht wird. Wir sind zuversichtlich, dass wir mit weiteren Hardwarekomponenten Standzeiten von 3-4 Monaten realisieren können, was sowohl den Wartungsturnus vor Ort als auch die Störungsanfälligkeit in sensiblen Flächen erheblich reduziert.

Durch eine zusätzliche Erweiterung der Laufzeit durch ein Solarpanel, welches den externen Zusatzakku speisen soll, wären prinzipiell noch längere Standzeiten denkbar. Durch den

geringeren Verbrauch beim ecoSNOOPER von 2-3Wh/Tag (im Gegensatz zu Systemen auf Raspberry Pi Basis, die bei vergleichbaren Anforderungen grob zwischen 14-38Wh/Tag verbrauchen würden) würde ein kleineres Solarpanel durchaus für den Tagesbedarf des ecoSNOOPERs ausreichen. Jedoch haben Systeme mit Solarpanel den Nachteil, dass sich diese nicht unbedingt mehr unscheinbar ausbringen lassen und so eine gewisse Sicherung vor Ort unter Umständen mit eingeplant werden muss. Wir erwarten, in der ersten Jahreshälfte 2024, eine lauffähige Version zur Verfügung zu haben.

## Notwendigkeit einer Vogel KI für Hessen

Auch wenn die Entwicklung von KI Systemen wie BirdNET ohne Zweifel beeindruckend ist und ganz neue Ansätze für Forschungsfragen und im Monitoring eröffnet, so bedarf es einer sorgfältigen Validierung und Nachbereitung um ökologische Information vom Datenrauschen falsch-positiver Nachweise zu befreien. Während die Diskussion bereits intensiv auf einige Aspekte eingeht, soll an dieser Stelle ein (praktischer) Ausblick gewagt werden, wie sich der datenanalytische Teil des Dauermonitorings zukünftig entwickeln ließe, um teil- oder gar vollautomatische Auswertungen zu ermöglichen. Seit der Version 2.4. erlaubt BirdNET das Trainieren eigener "Classifier". Diese bestehen aus gelabelten Aufnahmen, die je nach Classifier (z.B. Art) in die BirdNET interne Datenstruktur integriert/implementiert werden müssen. Aufgrund des bereits diskutierten Umstandes, dass BirdNET nicht nach innerartlichen Unterschieden von Lautäußerungen differenziert, sondern beispielsweise Gesang und Rufe zusammen in einem Classifier verwendet, erwarten wir erhebliche Verbesserungen, wenn Einzelarten nach Gesang und Rufen unterteilt als Classifier vorliegen würden. Der wachsende Datenbestand eines Dauermonitorings gepaart mit weiteren Datenquellen könnte bei entsprechendem Labelling eine brauchbare Datengrundlage liefern, um ein Hessen-spezifisches lokales KI-Modell für Vögel zu trainieren und fortlaufend zu verbessern. Neben der verbesserten Artbestimmung würden innerartliche Classifier auch zusätzliche ökologische Auswertungen erlauben (etwa zur genaueren Abgrenzung von Balzzeiten oder im Zugvogelmonitoring).

Eine weitere Möglichkeit zur Verbesserung automatischer Datenanalysen könnten zeitlich dynamische Artenlisten sein, die BirdNET für die Analyse verwendet oder im Nachgang zum Filtern der Ergebnisse herangezogen würden. Diese könnten sich entweder grob an Wertungsgrenzen aus den etablierten Methodenstandards orientieren, oder aber über ornitho-Daten gezogen werden, welche bekanntlich auch die zeitliche Dynamik bestens abbilden. Somit würden etwa Rotdrosseln im Sommer, entweder gar nicht erst von BirdNET berücksichtigt, wenn eine dynamische Artenliste in die KI gespeist würde oder aber pauschal im Nachgang ausgefiltert werden. Je nach Fragestellung wären die zumindest avifaunistisch

interessanten Ausnahmeerscheinungen jedoch "Opfer" eines solchen Filterprozesses. Würde aber ein Dauermonitoring eh den lokalen Artenbestand im Blick haben, würden gebietsfremde Arten eher eine untergeordnete Rolle spielen. Eine Implementierung von dynamischen Artlisten hängt somit stark vom Ziel des Monitorings ab.

# Literatur

- Darras K, Batáry P, Furnas BJ, Grass I, Mulyani YA & Tschardt T 2019: Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. *Ecological Applications* 29(6): e01954.
- Hill AP, Prince P, Piña Covarrubias E, Doncaster CP, Snaddon JL & Rogers A 2018: AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution* 9(5): 1199–1211.
- Höchst J, Bellafkir H, Lampe P, Vogelbacher M, Mühlhng M, Schneider D, Lindner K, Rösner S, Schabo DG, Farwig N & Freisleben B 2022: Bird@Edge: Bird Species Recognition at the Edge. In: Koulali, MA., Mezini, M. (eds) *Networked Systems. NETYS 2022. Lecture Notes in Computer Science* 13464: 69–86. Springer, Cham.
- Kahl S, Wood CM, Eibl M & Klinck H 2021: BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics* 61: 101-236.
- Markova-Nenova N, Engler JO, Cord AF & Wätzold F 2023: Will passive acoustic monitoring make result-based payments more attractive? A cost comparison with human observation for farmland bird monitoring. *Conservation Science and Practice* 5(9): e13003.
- Pérez-Granados C 2023: BirdNET: applications, performance, pitfalls and future opportunities. *Ibis* 165(3): 1068-1075.
- Pettorelli N, Laurance WF, O'Brien TG, Wegmann M, Nagendra H & Turner W 2014: Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology* 51(4): 839–848.
- Robinson JM, Harrison PA, Mavoá S & Breed MF 2022: Existing and emerging uses of drones in restoration ecology. *Methods in Ecology Evolution* 13(9): 1899–1911.
- Sethi SS, Ewers RM, Jones NS, Orme CDL & Picinali L 2018: Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device. *Methods in Ecology Evolution* 9(12): 2383–2387.
- Sethi SS, Fossoy F, Cretois B & Rosten CM 2021: Management relevant applications of acoustic monitoring for Norwegian nature - The Sound of Norway. NINA Report 2064. Norwegian Institute for Nature Research.
- Singer D, Hagge J, Kamp J, Hondong H & Schuldt A 2023: Aggregated time-series features boost species-specific differentiation of true and false positives in passive acoustic monitoring of bird assemblages. 10.13140/RG.2.2.10556.82568.
- Wägele JW, Bodesheim P, Boulat SJ, Denzler J, Diepenbroek M, Fonseca V, Frommolt KH, Geiger MF, Gemeinholzer B, Glöckner FO, et al. 2022: Towards a multisensor station for automated biodiversity monitoring. *Basic and Applied Ecology* 59: 105-138.

# Anhang 1

Abb. S1: Netzwerk von Begleitarten für A) falsch-positive und B) echt-positive Detektionen hervorgehoben (in rot) beim Baumpieper (S. 35) sowie bei der Amsel (S. 36). Stärke der Verbindung zeigt proportional die Häufigkeit der gemeinsamen Detektionen an. Es wurden nur Verbindungen zu Begleitarten dargestellt, die mehr als fünf mal gemeinsam detektiert wurden.

*Fig. S1: Network of accompanying species for A) false-positive and B) true-positive detections highlighted (in red) for the Tree Pipit (p. 35) and the Eurasian Blackbird (p. 36). The strength of the connection indicates proportionally the frequency of joint detections. Only connections to accompanying species that were detected together more than five times are shown.*





## Impressum

Hessisches Landesamt für Naturschutz, Umwelt und Geologie  
Abteilung Naturschutz  
Europastr. 10, 35394 Gießen

Tel.: 0641 / 200095 58  
Fax: 0641 / 200095 62

Web: [www.hlnug.de](http://www.hlnug.de)  
Twitter: [https://twitter.com/hlnug\\_hessen](https://twitter.com/hlnug_hessen)

E-Mail Dezernat N3: [vogelschutzwarte@hlnug.hessen.de](mailto:vogelschutzwarte@hlnug.hessen.de)

Nachdruck - auch auszugsweise - nur mit schriftlicher Genehmigung des HLNUG

### **Ansprechpartner Dezernat N3, Vogelschutzwarte**

Dr. Simon Thorn 0641 / 200095 38  
*Dezernatsleitung*

Lisa Eichler 0641 / 200095 36