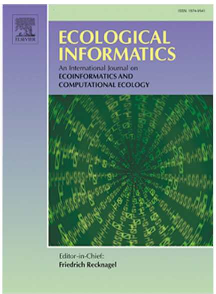


Territorial Acoustic Species Estimation using acoustic sensor networks

Leonhard Brüggemann, Daniel Otten, Frederik Sachser,
Nils Aschenbruck

PII: S1574-9541(25)00290-0
DOI: https://doi.org/10.1016/j.ecoinf.2025.103281
Reference: ECOINF 103281

To appear in: *Ecological Informatics*

Received date : 8 January 2025
Revised date : 12 June 2025
Accepted date : 12 June 2025

Please cite this article as: L. Brüggemann, D. Otten, F. Sachser et al., Territorial Acoustic Species Estimation using acoustic sensor networks. *Ecological Informatics* (2025), doi: https://doi.org/10.1016/j.ecoinf.2025.103281.

inst1a inst2b

Highlights

Territorial Acoustic Species Estimation using Acoustic Sensor Networks

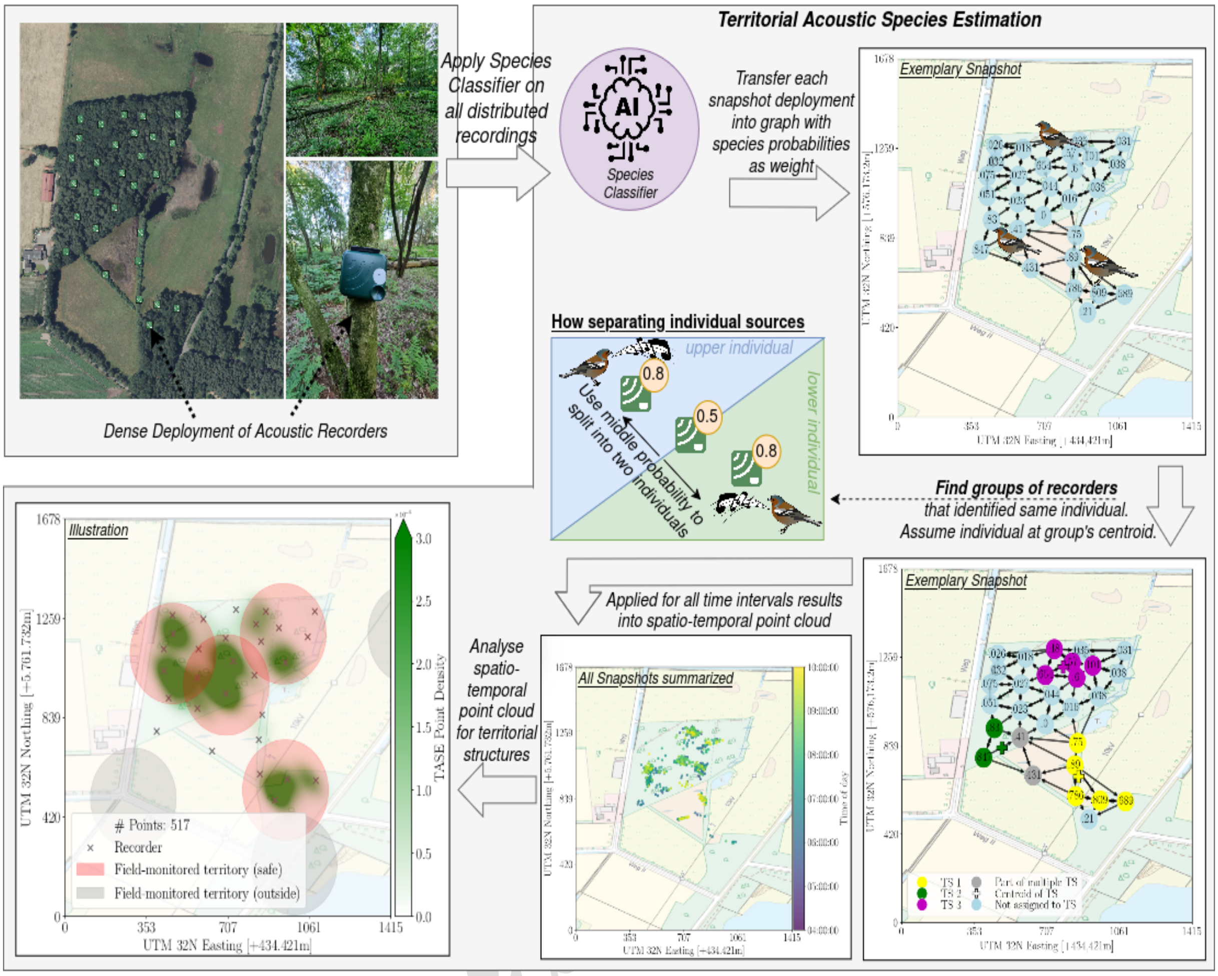
Leonhard Brüggemann, Daniel Otten, Frederik Sachser, Nils Aschenbruck

- Development of TASE: We formalize an automated Territorial Acoustic Species Estimation algorithm called TASE.
- Application of TASE on birds: We apply TASE on birds using a state-of-the-art acoustic species classifier.
- Proof-of-Concept-Evaluation in bird acoustics: We apply TASE in a 12 hectare real-world deployment and prove that it works on par with expert's field monitoring methods.
- Publication of TASE-ASNet: We publish and share the first cohesive acoustic dataset from our deployment, including field monitoring results, as Open Data.

Graphical Abstract

Territorial Acoustic Species Estimation using Acoustic Sensor Networks

Leonhard Brüggemann, Daniel Otten, Frederik Sachser, Nils Aschenbruck



Territorial Acoustic Species Estimation using Acoustic Sensor Networks

Leonhard Brüggemann[a], Daniel Otten[a], Frederik Sachser[b], Nils Aschenbruck[a]

[a] *Osnabrück University - Institute of Computer Science, Friedrich-Janssen-Str. 1, Osnabrück, 49076, Germany*
[b] *Austrian Research Centre for Forests, Seckendorff-Gudent-Weg 8, Wien, 1130, Austria*

Abstract

Accurate biodiversity assessment is fundamental for effective conservation management and environmental policy-making. However, monitoring local species populations is time-consuming, as experts can cover only one limited area at a time and are also prone to errors due to their varying knowledge and experience. Advances in low-cost autonomous recording units and AI-based classifiers offer new tools for species monitoring. However, while helpful in identifying species, current tools for acoustic species monitoring fall short in providing data on local populations. This limitation emphasizes the demand for more sophisticated methods, as uncertainties in estimating species populations can lead to misleading conclusions and misclassification of conservation statuses. In this work, we take a significant step towards more sophisticated monitoring by presenting a Territorial Acoustic Species Estimation approach, called TASE, to extract spatial, territorial patterns of species using acoustic sensor networks, allowing the estimation of territorial individuals of a species. It requires a distributed sensor network and exploits the characteristic spatial distribution of territorial species. We formalize TASE, apply it to bird acoustics, and share a proof-of-concept evaluation in a real-world deployment in a nature reserve, deploying 29 devices over 12 hectares. We show that it works on par with the time-consuming practice applied by bird experts and can provide novel insights into the spatial use of sound-producing territorial species.

Keywords: acoustic monitoring, species abundance, birds, BirdNET, AudioMoth, TASE

1. Introduction

Accurate information about local species populations is crucial for assessing species' local survival and also serves as an early warning system for environmental and ecosystem health in a given region. This knowledge can lead us to take proactive measures for species survival and environmental conservation, while uncertainties can lead to deceptive conclusions or improperly classifying species' conservation statuses. Acquiring data about species populations is time-consuming as each site needs to be visited multiple times by experts and is linked to various issues, ranging from disturbances of the species to methodological bias due to varying knowledge and experience (e.g., [43, pp. 67-69] [49, pp. 26-68]).

Passive acoustic monitoring (PAM) offers new insights into ecological questions by deploying autonomous acoustic recorders (ARUs) in natural environments [42]. Tools can effectively survey sound-producing species and transform monitoring practices. Unlike human observers, PAM continuously collects field data without any disturbance of the environment [1]. While manual data analysis is impractical due to the vast amounts of data, recent advances in classification models, like the Open-Source classifier BirdNET[27] for acoustic bird species identification, provide automated processing capabilities. Today, we are reaching more and more a state where acoustic monitoring evolves into a suitable complement or alternative to traditional field surveys [38].

Existing classifiers are effective at detecting species and inferring occupancy status and are extensively used in nature conservation, ecological sciences, and monitoring programs worldwide. However, ecologists urgently search for species abundances, i.e., to count the individuals of each species at a recording location or multiple recording locations. To this date, no automated approaches have proven their application in practice. Existing studies focus on identifying individuals in recordings based on unique acoustic features, a process known as Individual Acoustic Monitoring (IAM). This involves extracting individual features from acoustic signals, organizing them in a feature space, and identifying clusters to estimate the number of individuals. For various species, manual or semi-automatic methods can extract individual traits from vocalizations [28]. However, research often targets a single species, as in [17, 35] using manual methods, apply techniques like Gaussian mixture models, hidden Markov models [2, 40], support vector machines [12], or neural networks [18]. While focusing on specific species can improve accuracy, it limits broader adoption and requires method development or adaptation for each new species or environment. Some multi-species approaches, such as [48], rely on expert-driven steps, like separating foreground and background sounds, which are challenging in noisy environments with overlapping vocalizations. According to [28], IAM performs poorly under complex conditions and mainly focuses on species with simple vocalizations. Environmental factors, including wind, water, and overlapping sounds, mask individual vocalizations, reducing transferability and generalizability [28]. Key challenges include determining which acoustic features best identify individuals and developing methods without expert input. Finally, without individual identification, automated localization of individuals remains infeasible [41]. Above's limitations on IAM does not apply to methods that rely on *species* classification. A promising approach by [33, 34] uses counts of species vocalizations to infer animal density. However, its application still relies on laborious expert validation of sound files.

Another approach to measure individuals is presented by [22, 47]. They assume a species classifier, meaning it can identify merely *species* in a recording, but not *individuals*, so they always assume it to be one organism. If multiple ARUs are deployed whose detection ranges overlap, they analyze the spatial coverage from which they derive the maximum number of simultaneously vocalizing individuals. Their approach has a critical challenge for real-world deployments: it is based on an apriori known detection range of the devices, which is, in practice, highly dynamic. Various factors affect it, such as the recording equipment, specifics, placement, interfering acoustics from other species, and environmental noise such as wind [50]. Furthermore, the algorithm presumes that all sources are consistently active within a given time window. However, birds avoid singing simultaneously, leading to undercounting as not all individuals are vocally active simultaneously, which is proven, e.g., for birds [14]. While their algorithm is not applicable in natural environments, it offers an intriguing approach by connecting to advancements in automated species classification and affordable ARUs, which can form an acoustic sensor network (ASN).

This paper presents TASE (Territorial Acoustic Species Extraction), an algorithm designed to estimate species populations in acoustic monitoring. Given an ASN that gathers data from a sizeable cohesive area, we focus on animal species that produce sound to exclude conspecifics from their territory, creating peculiar spatial patterns. By applying a state-of-the-art species classifier on every node in our ASN, we derive the same peculiar spatial pattern per species, the territories. Given that, it is possible to quantify positive or negative trends in animal populations. That is especially true for songbirds because the number of territories equals the number of males and can be used as a proxy for animal abundance. Such an ASN gathers huge amounts of data, making automated processing mandatory. Our main contributions are as follows:

- Development of TASE: We formalize an automated Territorial Acoustic Species Estimation algorithm called TASE.
- Application of TASE on birds: We use a state-of-the-art acoustic bird species classifier.
- Proof-of-Concept-Evaluation in bird acoustics: We apply TASE in a 12-hectare real-world deployment and prove that it works on par with expert field monitoring methods.
- Publication of TASE-ASNet: We publish and share our deployment's first cohesive acoustic dataset, including field monitoring results, as Open Data.

2. Related Work

To the best of our knowledge, no approaches in computer science follow our approach of capturing species' territorial spatial patterns based on their acoustic sounds as a proxy for measuring species abundance. Our algorithm leverages species territoriality. Thus, we provide an overview of the territoriality of sound-producing species to understand the biological aspect of our approach better.

Territorial Behavior of Species

Our approach incorporates species' territorial behavior, so we first provide some background. On a broader scale, many factors, including the quality and availability of habitat, connectivity, competition, and behavioral characteristics of the target species, form the distribution and occupancy patterns of organisms. An individual's home range describes the whole area that is regularly used to fulfill all its requirements, whereas a territory is a defended area within the home range.

Territorial behavior in birds is well-studied [36], ranging from small nesting territories in colonial species (e.g., Barn Swallow *Hirundo rustica*) to feeding territories maintained by European Robins *Erithacus rubecula* in winter. Many songbirds defend territories during the breeding season for mating, nesting, and feeding, using species-specific songs to repel conspecifics. Territoriality has been researched for over a century [25], aiding abundance estimates through expert field monitoring. Although ARUs now facilitate studies on phenology, species composition, and large-scale distributions [13, 11, 44], their focus remains on species identification rather than estimating species abundance.

In territorial species, individuals spatially segregate and defend areas against conspecifics to secure mates or exclusive resources [39]. Their vocal interactions often reflect the spatial arrangement of territories and potential breeding pairs, making them valuable for bioacoustic analysis. Territoriality has been examined in nearly all vertebrates [31], though its definitions vary [32]. Here, we focus on sound-producing organisms that acoustically assert exclusive space use. For our approach, the following territorial characteristics are key:

- Distribution of territories is dynamic: In detail, the territoriality of a species is not a simple but a complex dynamical phenomenon. The territory's boundaries are not necessarily clear and can overlap, resulting in interactions on an individual level [26, pp. 716-718]. Additionally, individuals might die or lose their territory as conspecifics take over.

- Territories differ in sizes: Individuals of the same species occupy territories of different sizes, as these vary, for example, with habitat quality and conspecific density [16]. Consequently, population densities of species differ, and their transferability to other areas is limited. Due to a cap on expenses, traditional monitoring programs rely on data from only a few site visits. The density or abundance of a species needs to be calculated without precise information on territorial extends, thus indirectly assuming a uniform distribution of equal-sized territories.

- Territories are not constantly proclaimed: Not all species show territorial behavior constantly, e.g., many birds only during mating season [6, pp.274-275]. Many species show species-specific vocalization activity patterns, leading to individuals proclaiming their territory only seasonally or at different daytimes, e.g., in the morning and evening [46].

3. Formalization

Before defining the problem our approach solves, we outline the real-world deployment of this work. We then briefly examine the dataset and eventually formalize the challenge.

3.1. Deployment

A large-scale deployment of 29 ARU (AudioMoth v1.2) devices was conducted on June 3, 2023, from 4:00 to 10:00, resulting in six continuous soundscapes (174 hours total). The deployment occurred in a 12-hectare nature reserve in North Rhine-Westphalia, Germany, home to approximately 35 bird species, alongside human-made sounds from a nearby street and farm. The area features diverse habitats, including forests, hedges, ponds, heather fields, and meadows, creating a complex acoustic environment with potential interference from mammals and insects.

(a) Deployment area, the icons refer to the ARUs

(b) Exemplary Scene within the forest

(c) ARU in its waterproof case

Figure 1: Deployment site and recording devices

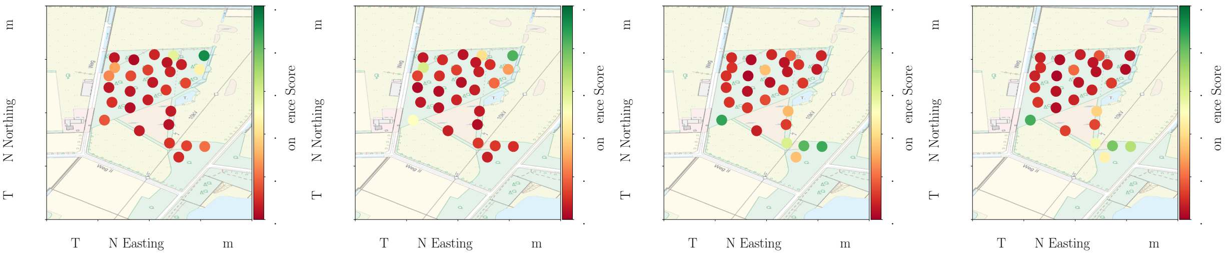
Based on preliminary tests and insights from bird experts, ARUs were strategically positioned very densely - typically just 25 to 50 meters apart within forested areas and no more than 100 meters apart in open spaces. The intentional closure and atypical spacing of recording units, e.g., when focusing on detecting species presence, were deliberately implemented to ensure substantial overlap between adjacent units, thereby providing continuous acoustic coverage essential to our approach. We synchronized all recorders with the AudioMoth app. Although the clocks still experience minor drift, 12-hour bench tests show a maximum deviation of about 400 ms, which is negligible for our purposes. Because we do not attempt fine-scale localization, but instead make a crude estimation based on every node whose 3-second classification windows overlap, this drift remains comfortably inside that 3-second window.

The devices were configured with firmware version 1.7.1, a 48kHz sampling frequency, 16-bit depth, microphone gain of 4, and a recording duration of 3595s with a 5s sleep interval. Micro-SDs were collected at the end of the deployment.

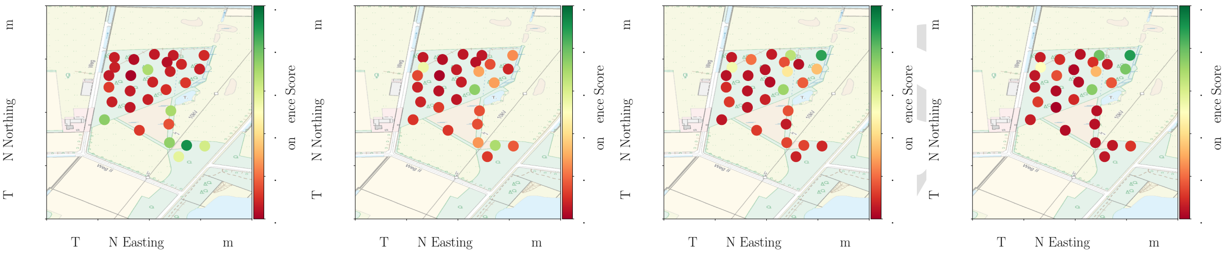
3.2. Problem Definition

Before formalizing the problem, our approach will solve some background. When deploying a set of nodes, we acquire a set of soundscapes that capture species' territorial vocalizations. Applying a species classifier to these recordings yields unitless classification scores (ranging from 0 to 1) for each identifiable species, e.g. [51], in the following referred to as *confidence scores*. A score of 0 indicates the species is not detected, while a score of 1 indicates a strong species detection. Across distributed nodes in an ASN, this generates spatially distributed confidence scores for each species. As the species vocalizes, its signal propagates to neighboring nodes, causing peculiar spatial patterns in the confidence scores. These confidence scores comprise the territoriality of sound-producing species. Due to territoriality, species remain within a specific area. Over time, these areas consistently have higher confidence scores, indicating species territories that serve as effective approximations for species abundance.

Fig. 2 illustrates confidence score sets over time using bird data from a real deployment, representing a small subset of our dataset (see Sec. 3.1 for details). Each circle represents a recording node, colored by the classifier's confidence score of identifying the Common Redstart *Phoenicurus phoenicurus*. The recordings were analyzed with BirdNET v2.4 [27], providing confidence scores for each three-second window over ten seconds. During this time, confidence scores fluctuate with bird



(a) Time: 4:15:00 - 4:15:03 (b) Time: 4:15:01 - 4:15:04 (c) Time: 4:15:02 - 4:15:05 (d) Time: 4:15:03 - 4:15:06



(e) Time: 4:15:04 - 4:15:07 (f) Time: 4:15:05 - 4:15:08 (g) Time: 4:15:06 - 4:15:09 (h) Time: 4:15:07 - 4:15:10

Figure 2: Schematic of a real ASN for birds. BirdNET identifies Common Redstart *Phoenicurus phoenicurus* on June 3, 2023, using a sensitivity of 1.5, 2-second overlap, and a minimum threshold of 0.0.

activity. One bird is active in Figs. 2a and 2b in the top right corner, resulting in high confidence scores. Then, it is silent in Figs. 2c to 2f, and active again in Figs. 2g and 2h. Similarly, another bird in the bottom right is initially inactive, vocalizes, and then becomes silent. A third Common Redstart in the west causes high confidence scores at neighboring nodes. Some nodes, such as the green node at the top in Fig. 2b, show high confidence scores while their neighbors do not, potentially indicating false positives. This is due to the inherent difficulty of acoustic species classification, where models may mistakenly detect patterns that resemble a species even when it's not present. This issue arises from classification errors inherent in the AI model. Therefore, it is crucial to ensure that subsequent processing steps of finding territories mitigate these errors to prevent inaccurate results. Our evaluation (see Sec. 6) demonstrates that our algorithm effectively copes with these classification errors.

Formal Definitions

After the informal introduction of the problem we will solve in this paper, we now provide a formal definition of the problem. The classifier comprises several parameters denoted as c. Overall, the classifier is a function that takes in a sound sample beginning at t_{start} and ending at t_{end} and outputs a vector of confidence scores. The vector consists of s entries, each referring to the presence of a particular species. In short, the classifier can be written as

$$F_c \colon [t_{\text{start}}, t_{\text{end}}] \to [0,1]^s. \tag{1}$$

Obtaining the confidence scores that a certain species is present at a given time point m in the audio necessitates a shifting window. Let $w \in \mathbb{N}$ be the size of that time window, with the time window being $W_m := [m, m+w]$ where $m \in \{0, 1, \ldots, d\}$ and d being the length of the audio. Moving the time window m in the audio gives a set of sliding windows $W := \{W_0, \ldots, W_{d-w}\}$. Note that windows with a starting point $m > d - w$ exceed the audio duration d. By sliding the time interval over the full audio duration, we can identify the presence of a species at a given point in the audio.

Given an acoustic sensor deployment, we have a set of recorders (nodes) $V = \{v_0, v_1, \ldots, v_n\}$ with well-defined locations and well-synchronized timing, ensuring time-synchronized recordings. For simplicity, we assume continuous acoustic recordings and equal recording duration d for all nodes.

Given this set of nodes $V = \{v_0, \ldots, v_n\}$ and a synchronized time interval $[t_i, t_j]$, we define a function G that applies the classifier F_c to all nodes in V over this time interval:

$$G_c \colon V \times W_m \to \{F_c(W_m)\}^n = [0,1]^{s \times n} \tag{2}$$

Using this function on every time window yields the following data:

$$G_c(V, W) = \bigcup_{[t_i,t_j] \in W} \bigcup_{v \in V} \{G_c(v, [t_i, t_j])\} \tag{3}$$

Now that we have formulated the input data, we still need to define the target of our problem. The signals processed by the classifier originate from different territorial individuals. Let $b \coloneqq \{b_1, \ldots, b_k\}$ denote that set of individuals. Given the input data $G_c(V, W)$, the target is to obtain the number of territories k.

4. Territorial Acoustic Species Estimation (TASE)

Our Territorial Acoustic Species Estimator, called TASE, aims to solve this challenge. We substantiate our requirements, formalize the workflow, and discuss its complexity.

4.1. Requirements

The algorithm is based on the assumptions stated below. These simplifications are necessary when developing this cross-domain algorithm incorporating knowledge from species-related research (see Sec. 2) in which territorial individuals might show complex behavior that is difficult to model.

Time-synchronized Recordings: As defined, we require a deployment in which all nodes record their acoustic surroundings simultaneously, assuring that any post-processing refers to the same moment. That can be fulfilled through a base station connecting and acquiring data from all nodes. The node's precise synchronization can be assured, e.g., by GPS[1].

Classifier's confidence score relates to exactly one source: As already stated, automated *individual* identification is not possible yet (see Sec. 1), and acoustic species classifiers are only capable of detecting a *species*. Thus, we assume that one individual causes a classifier's confidence score of a species.

Distant-dependent decline of Classifier's confidence scores: When a sound source emits a signal, the node nearest to the source has higher classifier confidence scores than more distant nodes. This was shown by [15, 37]. For simplicity, we assume that G_c, the input set containing all species confidence scores for all nodes is monotonically decreasing with increasing Euclidean distance between the sound source b_i and the node $v \in V$. With increasing distance, the species confidence score declines, and it becomes more likely to produce false positives, increasing the risk of inducing errors in further processing.

Node placement: At least two devices record each possible source location in the deployment area (so-called 2-coverage). Thus, the node placement is denser than the territories' distribution, resulting in nodes between two territories.

4.2. Concept

All requirements ensure that there is always an intermediate node between two territories. Due to the distance-dependent decrease in classifier confidence scores, nodes closest to a vocalizing source have higher confidence scores than intermediate nodes. Consequently, if neighboring nodes have higher confidence scores than an intermediate node, we can spatially separate these nodes into distinct groups corresponding to individual vocal sources. This pattern enables effective identification and separation of simultaneous vocalizations. By intersecting classification data from multiple recording devices over long periods, we identify nodes consistently recording the same individual staying in specific areas. Thus, we acquire a spatio-temporal point cloud that effectively captures territorial spatial patterns.

---

[1] https://s.gwdg.de/wjuNOL, last access: 17th May. '25

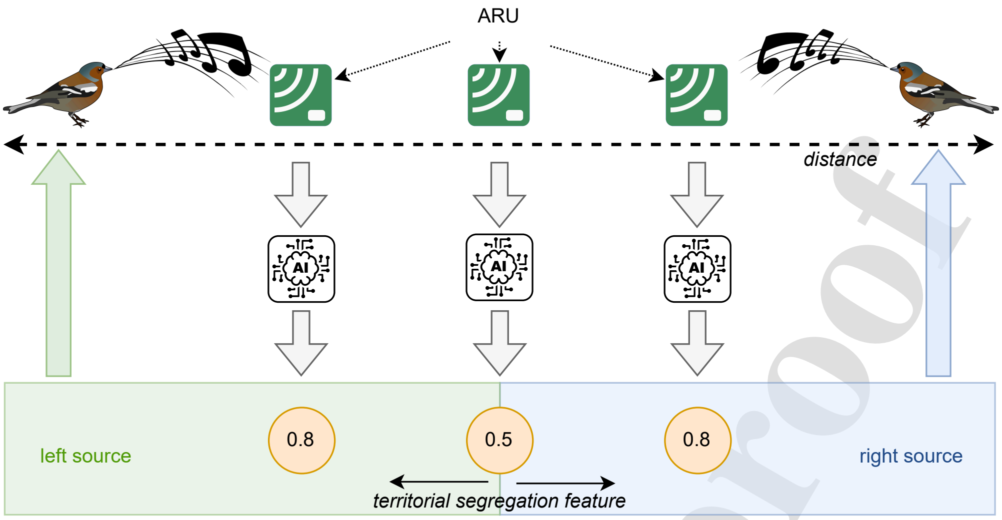
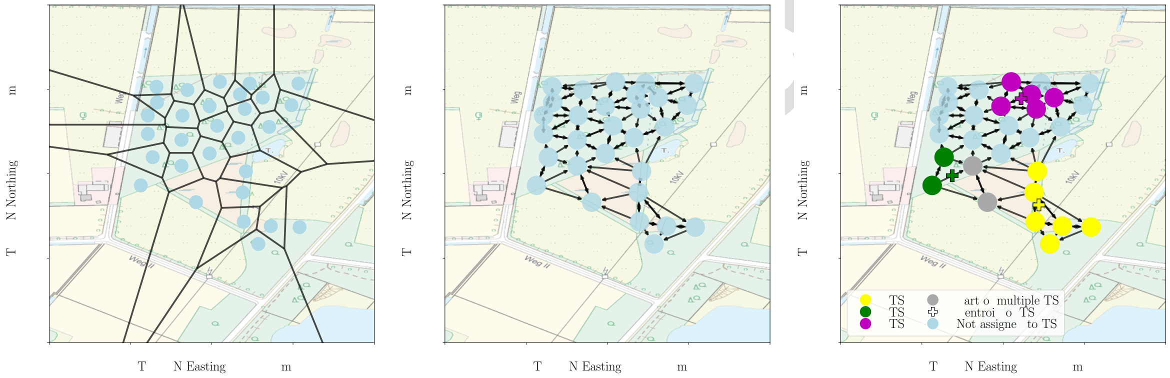


Figure 3: Illustration of how to separate simultaneously species vocalizations



(a) Voronoi diagram of ARU (b) Triangulated, weighted, directed graph (c) TS including their weighted centroids

Figure 4: Illustration of TASE's steps step 1 to 4 from the real-world deployment (see Sec. 5.2). Weights are without preceding zero.

4.3. Algorithm

TASE comprises five steps. First, we transform the sensor network into a graph. Second, we assign weights to the nodes. In the third step, the graph is divided into subgraphs referring to individuals. The first three steps are depicted in Fig. 4. The steps 2 and 3 are repeated for each classification period, and lastly, in steps four and five, merged to deduce frequently occupied areas and territories (exemplarily visualized in Fig. 5).

4.3.1. Transfer the Deployment into a Graph

At first, the deployment area is split into regions closest to each node. This is done by creating a Voronoi diagram, also known as Dirichlet tessellation or Thiessen polygons [30]. As the classifier is monotonously decreasing with respect to the distance, we can assign a sound source directly to a cell. As the source's signal propagates into neighboring Voronoi cells, other nodes will record and identify it. To model this relation, we connect two nodes via an edge if the corresponding Voronoi cells are neighboring. If the Voronoi cell of two nodes share a common border but the nodes are unreasonably far away, we remove this edge. This comes to hand if the distance is bigger than a species' maximum known territory size.

More formally, we build the directed Delaunay-triangulated graph $G(V, E)$ where $V = \{v_1, ..., v_n\}$ is the set of nodes. The nodes are projected into a 2D plane, e.g., an UTM projected coordinate system, so each node v_i has location p_i. Given a set of distinct points $P = \{p_1, p_2, \dots, p_n\}$ in a Euclidean plane, called *sites*, the Voronoi cell $Vor(p_i)$ corresponding to a site p_i is the set of all points x in the plane such that the distance from x to p_i is less than or equal to the distance from x to any other site p_j (for all $j \neq i$). Let $d(x, p_i)$ denote the Euclidean distance between the point x and the site p_i. The set of edges E contains all edges between adjacent Voronoi cells: $\forall e = (u, v) \in E,$ if $u, v \in V$ and $Vor(u)$ is adjacent to $Vor(v)$ and $d(u, v) \leq d_{max}$ with d_{max} being a maximum reasonable distance.

4.3.2. Apply Weights to V and add Direction to Edges in E

For clarity, we focus on one species s and one-time window W_m throughout steps two to three. After generating the Delaunay-triangulated graph, we add the classifier's confidence score of time window W_m to each node. Furthermore, we indicate the direction in which confidence scores between nodes decline by assigning a direction to each edge. An edge between two nodes always points from the node with a higher confidence score to the node with a lower confidence score. Formally, $\forall e = (v_i, v_j) \in E$, the edge points from v_i to v_j if and only if $w(v_i) \geq w(v_j)$.

The core idea behind this can be summarized as follows: Given that two sources emit a sound simultaneously at some distance, the recording nodes between them should have only incoming edges and no outgoing edges, thus separating both sources. We denote this characteristic as *territorial segregation feature* (see Fig. 3): Given the weighted, directed Delaunay-triangulated graph $G = (V, E)$, sources s_0 and s_1 of the same species, and a confidence score function F_c that is monotonically decreasing with distance, a node $\bar{v} \in V$ where $d(\bar{v}, s_0) \approx d(\bar{v}, s_1)$ will have no or a low number outgoing edges. Outgoing edges might occur, e.g., if neighboring nodes have identical confidence scores.

4.3.3. Detecting Territory Subgraphs

By design, the underlying directed graph with its nodes' weights can exploit the *territorial segregation feature*. Connected areas in the graph belonging to the same individual are referred to as *Territory Subgraphs (TS)*. Formally, they are constructed as follows:

First, the node with the highest detection confidence score is marked as the root node of the territory. Using a breadth-first search (BFS), we add neighboring nodes whose confidence score is less and likely caused by the same sound source. The BFS ends if and only if a node with only incoming edges is reached. These *border nodes* mark the boundary between two TS. Once the BFS terminates, all nodes and edges added to the TS are removed, except for the border nodes. The procedure is then repeated for the nodes not included in the previous TS, starting with the node with the remaining highest confidence score, followed by the next highest, and so on. Eventually, the graph is split into multiple TS, each representing a different individual. Some nodes, however, remain unassigned because they do not meet the root-node criteria or the criteria to be part of a TS. Detecting all TS can be efficiently implemented by first sorting the root nodes R according to their weights. Then, we perform BFS and remove all nodes from G and R that are part of the TS. To cope with things like misclassifications, we added some extra rules.

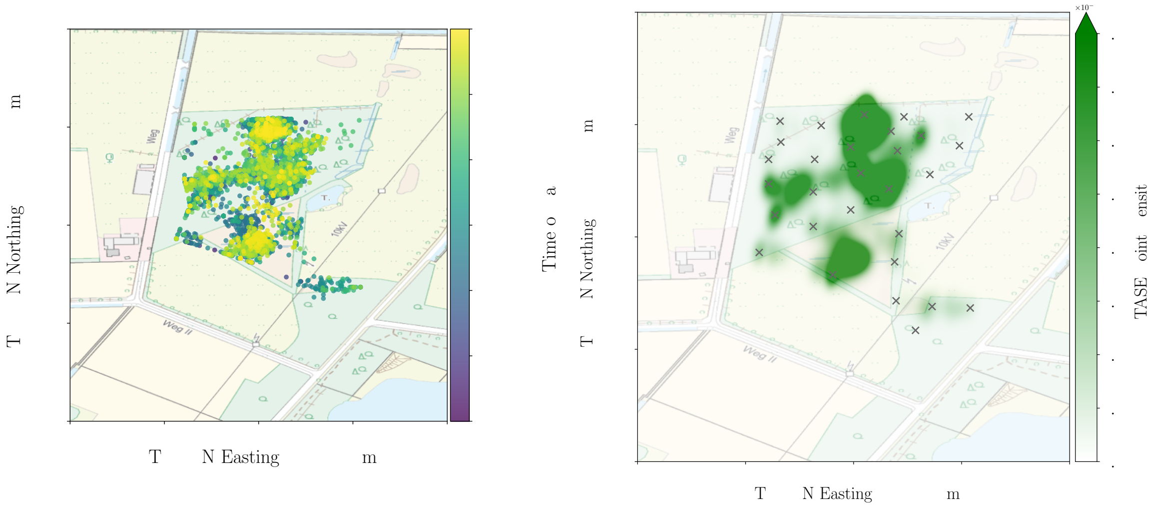
Threshold for a subgraphs root nodes R: The higher the classifier's confidence score, the closer the distance to the sources, and the less likely false positives are. Thus, we define a threshold threshold_R for all nodes $v \in R$ such that $w(v) \geq \text{threshold}_R$. In other words, if the confidence score provided by a certain node is below that threshold, it is not selected as a root node.

Threshold for a cluster's border-node: As with increasing distance the likelihood of false positives increase, also a threshold for the border-nodes is set. We define $threshold_B$ for $\forall v \in V, deg(v) \approx 0 : w(v) \geq threshold_B$.

Set maximum distance from root to border-nodes: We define a maximum distance from a node $r \in R$ to a border-node, depending on the territory size of the species. This criterion can be formalized as $\forall v \in V \setminus \{r\}, \quad d(r, v) \leq threshold_T$.

4.3.4. Derive Representatives for Subgraphs belonging to territorial Individuals

Due to the local movement of the territorial individuals, the TS differs for each time window W_m. Therefore, we need to merge the information from all TS to obtain an overview of the territories. To



(a) Step 4: Calculate spatio-temporal point cloud

(b) Step 5: Analyze density pattern for territoriality

Figure 5: Exemplary visualization of TASE step 4 and 5 referring to species Blackcap *Sylvia atricapilla* in the deployment (see Sec. 3.1)

achieve this, a representative location for each subgraph is calculated by the centroid of the nodes' locations weighted by their classifier's confidence score. Eventually, we acquire a spatio-temporal point cloud, with high point densities in the territories' areas. It can be visualized, e.g., by a kernel density estimation.

4.4. Complexity

The complexity of the algorithm is made up by the complexity of every single step. The first step consists of the Delaunay Triangulation for a set of nodes $V = \{v_0, \ldots, v_n\}$ to get the edges E between adjacent Voronoi cells. The complexity of this step is $\mathcal{O}(Vlog(V))$ [30]. Second, the weights are applied, and the graph is directed. This can be done in $\mathcal{O}(E)$. Finding the subgraphs root nodes takes $\mathcal{O}(Vlog(V))$, the BFS needs $\mathcal{O}(V + E)$ steps, and removing nodes from a territorial subgraph from the root nodes requires $\mathcal{O}(V)$. At most the whole graph belongs to one territory subgraph. Thus, finding the center is done in $\mathcal{O}(V)$. Applying above to all time windows W, step one to four have a complexity of $\mathcal{O}(Vlog(V)) + \mathcal{O}(W \cdot (E + Vlog(V)))$. As our algorithm uses a Delauny-triangulated graph, which is a specific kind of planar graph, we know that $E = \mathcal{O}(V)$. By substituting the second term, we can simplify to $\mathcal{O}(Vlog(V)) + \mathcal{O}(W \cdot Vlog(V))) \approx \mathcal{O}(W \cdot V \cdot log(V))$.

5. Applying TASE on Birds

We presented the general approach, which is now applied to birds due to their territorial behavior and sound-producing capabilities. Furthermore, they are well-researched species, acoustically identifiable by today's AI-based species classifiers, and serve as good indicators of environmental health and ecosystem changes, existing almost everywhere on Earth (e.g., [19]). However, when applying TASE to birds, additional challenges arise from the increased presence of overlapping sounds, such as calls from multiple individuals or background noise from other species. These interferences are more pronounced in avian soundscapes, particularly during events like the dawn chorus, where vocalizations overlap frequently. Consequently, we will adapt our algorithm to address these complexities, making it well-suited for avian soundscapes.

Due to its consistent and reliable performance [21, 29], as well as its Open-Source availability, we have chosen to focus on the BirdNET classifier [27]. BirdNET is applied on the recordings from the deployment described in Sec. 3.1. Note that ARUs were not acoustically calibrated, for two reasons.

First, environmental variability is an inherent feature of real-world data that the model should accommodate rather than eliminate. Second, calibrating dozens of recorders in remote, heterogeneous habitats is logistically unrealistic and site-specific. To mitigate the resulting variation in BirdNET confidence scores, we introduce the $threshold_\delta$ in the next section. However, our approach does not rely solely on this threshold. The ARUs are densely spaced and record continuously, so an erroneous classification on one unit is usually countered by the threshold_R for the root-nodes and the higher-confidence detections from neighbouring units when a bird vocalizes. This redundancy keeps overall detection reliability high, as confirmed by our results.

TASE requires, as stated in Sec. 4.1, the species classifier requires its confidence scores to correspond to the distance between a recording device and a sound source in a monotone relation. In the following section, we examine whether this is true for BirdNET and assign values to the TASE parameter defined above.

Table 1: Correlation between distance and confidence

		sensitivity[1]				
species	reference_{db_A}	0.50	0.75	1.00	1.25	1.50
T. philomelos	100	-0.47	-0.48	-0.48	-0.49	-0.50
F. coelebs	92	-0.52	-0.53	-0.54	-0.55	-0.55
L. megarhynchos	86	-0.51	-0.52	-0.53	-0.53	-0.53
L. megarhynchos	75	-0.50	-0.51	-0.52	-0.53	-0.53
R. regulus	74	-0.70	-0.71	-0.73	-0.74	-0.75

[1] sensitivity refers to BirdNET's configuration parameter, not the machine-learning evaluation metric

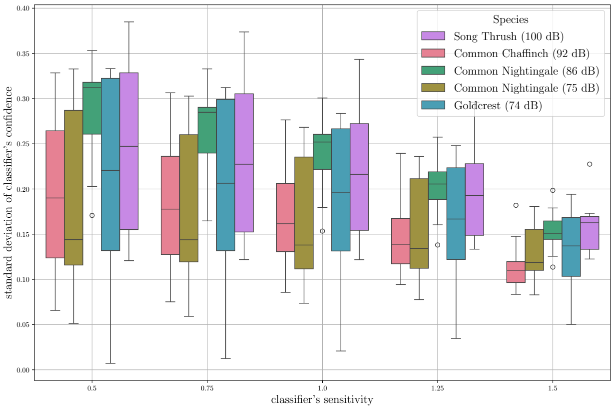


Figure 6: Standard deviation of BirdNET's confidences per species

5.1. *Distant-dependent Decline of BirdNET's Confidences*

A correlation has already been documented for bird classifiers in literature, e.g., [15], and also the classifier BirdNET we use [37]. However, the latter does not focus on European species but three American species, and it does not mention the BirdNET model v2.4 or consider different configurations. We wanted to verify this for typical European species and gain more insights into how the confidence scores change depending on different values for configuration parameter *sensitivity*. This parameter is crucial, affecting the gradient of sigmoid-scaled activity function [27] and is meanwhile known to have a great impact on the classifier's performance according to [20]. It has not been examined regarding the distance yet.

In a deployment from 16th to 19th December 2019, we acquired real acoustic data with an ARU Audiomoth v.1.2 [23]. We placed the ARU in a straight line from the speaker device at 1m and every 10m up to 100m. At each distance, we replayed four common European species in realistic amplitudes, referring to the measurements of [7, 9]. The audio chunks have manually been extracted and classified separately. As the recordings per distance have been made one after the other, the noise and interference vary due to human-made sounds in the far distance. Audio chunks with extremely high interference in which the species of interest could not be segmented manually were excluded, resulting in a different number of samples per distance.

Table 1 lists the correlation coefficients between BirdNET confidence and distance for all species. The correlations are moderate to strong, so confidence does not decline in a perfectly smooth fashion as distance grows. The negative values refer to that decline. Raising the sensitivity level increases the coefficients to 0.05, making the distance effect slightly clearer. This is likely caused by variations in confidence score values at different distances. To compensate for the classifier's variation, we modify TASE as follows:

1) After building the Delaunay-triangualted graph, we add bidirectional edges between nodes whose weights have a difference in confidence score of less than E_δ to each other. For $G = (V, E)$ the edges E become: $\forall e = (v_i, v_j) \in E$ e points from v_i to v_j if and only if $|w(v_i) - w(v_j)| \leq E_\delta$.

2) Given such a modified Delaunay-triangulated graph as above and an adverse distribution of weights, separating two simultaneous vocalizations becomes more erroneous. In order to prevent that, we have added new criteria. For a territorial subgraph $TS = (G_{TS}, E_{TS})$, the weights must decline or remain within a range from the root node to the border nodes. Every node $(u, v) \in E_{TS}$: $w(u) + TS_\delta \geq w(v)$ with TS_δ being a threshold for the confidence score difference.

In order to derive reasonable values for E_δ and TS_δ, we examined for our deployment the standard deviations of the species' confidences more closely. A significant, species-independent correlation between the distance and the confidence variations is not observed. Instead, we found a correlation between sensitivity and standard deviation, which is visualized in Fig. 6. With increasing sensitivity, the standard deviation declines notably. Given that observation, for different classifier sensitivities, the choice for E_δ and TS_δ differs, reaching about 0.2 for a sensitivity of 1.5.

Table 2: Summary of TASE' parameters in the evaluation

Classifier	
sensitivity	1.5
overlap	2
confidence threshold	0.0
Step 1: Building graph	
d_{max} [m]	100
Step 2: Add Weights and direction to graph	
E_δ	0.2
Step 3: Detecting territorial subgraphs	
$threshold_R$	0.5, 0.6, 0.7, 0.8
$threshold_B$	0.1
$threshold_T$	species-dependent
TS_δ	0.2
Step 4: Derive territorial subgraphs' representatives	
Method	weighted centroid

5.2. Parametrization

TASE uses a set of parameters across four steps to estimate territorial individuals, with two additional parameters for bird applications. Table 2 summarizes these parameters used in our evaluation in Sec. 6. However, we emphasize the urgent need for clear guidelines in acoustic monitoring, as much of the current knowledge is still based on practical experience rather than standardized protocols. Our parametrization should, therefore, be seen as a starting point.

Step 1: When building the Delaunay-triangulated graph, we set the maximum euclidean distance between two nodes d_{max} to 100 meters, because it is two times the recommended distance that bird experts cover [49, pp. 49].

Step 2: As shown in the previous section, we use a classifier's configuration with a sensitivity of 1.5 as it shows the highest correlation. Furthermore, we set E_δ and TS_δ to 0.2 based on our previous measurements.

Step 3: When identifying territorial subgraphs, we evaluate $threshold_R$ values of 0.5, 0.6, 0.7, and 0.8. Nodes exceeding $threshold_R$ are designated as roots of TS. For boundary nodes ($threshold_B$), we use BirdNet's default confidence threshold of 0.1. The maximum distance from a root to a boundary node ($threshold_T$) varies by species based on territory size and environment, requiring input from a species expert. We set $threshold_T$, the maximal distance between TS root to a border node, using the average bird density (breeding pairs per 10 hectares) from [4], assuming circular territories and defining $threshold_T$ equal to that radius.

Table 3: Comparison of the expert's ground truth and expert's interpretation of TASE' KDEs

Species	**Ground Truth**	**threshold$_\mathbf{R}$**				**Reference**
		0.5	0.6	0.7	0.8	
Anthus trivialis	1	1	1	1	1	Fig. 12
Certhia brachydactyla	4	4	4	4	4	Fig. A.1
Fringilla coelebs	4	4	4	4	4	Fig. A.2
Muscicapa striata	2 - 3	2-3	2-3	2-3	2-3	Fig. A.3
Phoenicurus phoenicurus	4 - 5	4-5	4-5	4-5	4-5	Fig. A.4
Phylloscopus collybita	3 - 4	3	3	3	3	Fig. A.5
Sylvia borin	1 - 2	1 - 2	1 - 2	1-2	1-2	Fig. A.6
Sylvia atricapilla	4	4-5	4-5	4-5	4-5	Fig. 7
Troglodytes troglodytes	2	2	2	2	2	Fig. 8
Erithacus rubecula	4 - 5	3 - 4[a]	3 - 4[a]	3- 4[a]	3- 4[a]	Fig. 10
Turdus philomelos	1	1[b]	1[b]	1[b]	1[b]	Fig. 11

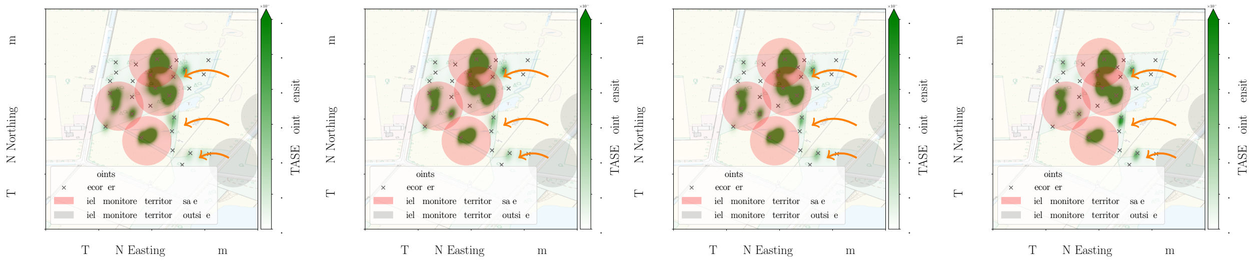
[a] Overlapping territories, [b] Extraordinary large territory

6. Evaluation

We apply TASE to a real-world dataset (cf. Sec. 3.1). The ground truth was provided by a bird expert who systematically traversed the field on the same day where the recording was taken, recording all observed birds on a map, following a methodology *similar* to the monitoring method known as territorial mapping [49, pp.47-59]. The main difference is that we carried out multiple field surveys within the same day over six hours rather than over a longer period. By repeatedly transecting the area — and given that the target species are territorial and highly vocal during the breeding season — this time-intensive approach allows us to identify stable spatial patterns that correspond to territory locations, even within a short timeframe. Similar methods have been applied, e.g., in [3]. In addition to the surveys conducted on the recording day, field surveys were also performed on prior days, which helped in the interpretation of the same-day observations. By understanding the spatial distribution of territorial individuals, the expert assesses the number of territories for that day, providing a ground truth. First, we qualitatively assess the similarities and differences between the spatial distributions identified by TASE and the expert's assessment. In the second step, we analyze cases where TASE performs poorly, highlighting its current limitations.

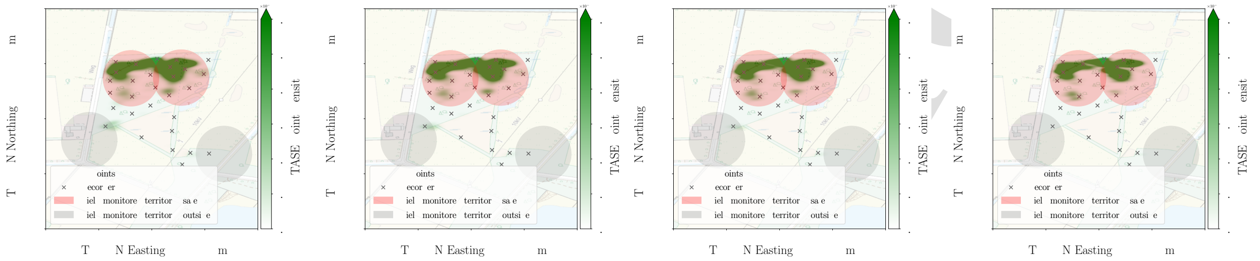
6.1. Comparing TASE to Expert's assessment

When applied to a species, TASE generates a spatio-temporal point cloud that implicitly captures the birds' sound-producing and territorial behavior. We anticipate the highest point density within territorial areas, representing the spatial extent and intensity of species activity. Kernel Density Estimation (KDE) is particularly well-suited for analyzing these patterns. It smooths the discrete points into a continuous density surface, highlighting areas of high intensity while preserving the spatial structure of territories. We applied KDE (with a bandwidth parameter of 0.15) to eleven species, using this as a baseline within our parameter space (cf. Sec. 5.2), and incorporated expert assessments represented as circles. These circles approximate the actual territorial areas, which are inherently dynamic and lack fixed boundaries (cf. Sec. 2).



(a) *threshold*$_R$: 0.5 (b) *threshold*$_R$: 0.6 (c) *threshold*$_R$: 0.7 (d) *threshold*$_R$: 0.8

Figure 7: Kernel density estimate (green) of spatio-temporal detections for the Blackcap *Sylvia atricapilla* generated by TASE, with the expert's territorial assessment shown by the colored circles.



(a) *threshold*$_R$: 0.5 (b) *threshold*$_R$: 0.6 (c) *threshold*$_R$: 0.7 (d) *threshold*$_R$: 0.8

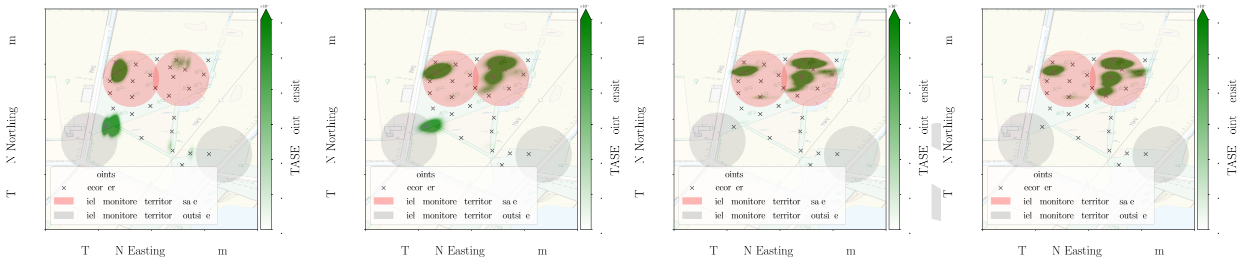
Figure 8: Kernel density estimate (green) of spatio-temporal detections for Eurasian Wren *Troglodytes troglodytes* after TASE, with the expert's territorial assessment shown by the colored circles.

We compare the expert assessment of territories based on field monitoring data and KDE analysis in Tab. 3. Assessing the number of territories from the KDEs involved a visual interpretation of the highest-density areas and their distance from each other. The expert analysis of KDE data focused on identifying high-density areas and evaluating their spatial distance. If the expert's assessment of KDEs fits exactly with the ground truth, we color the table's cell green. If it fits partly, we color it orange, and if not, we color it red. The initial analysis demonstrates that the proposed approach effectively aligns KDE assessments with ground-truth territory counts in most cases, underscoring the method's effectiveness. Although lowering *threshold*$_R$ predictably reduces the number of points, the overall pattern remains largely intact, yielding identical estimates across all species. The primary source of error appears to be overlapping territories, particularly for the species European Robin *Erithacus rubecula*, as the density maps reveal broad areas of activity rather than distinct, separated spots. However, this discrepancy was limited to one or two territories, confirming that the overall approach remains highly effective across most species and scenarios. From an ornithologist's perspective, missing a single territory is usually minor, especially in large populations. For rare or endangered species, however, it can significantly affect conservation decisions. Such miscounts are not unique to our method and are also common in traditional surveys. Fortunately, this issue does not apply to the species considered in our study.

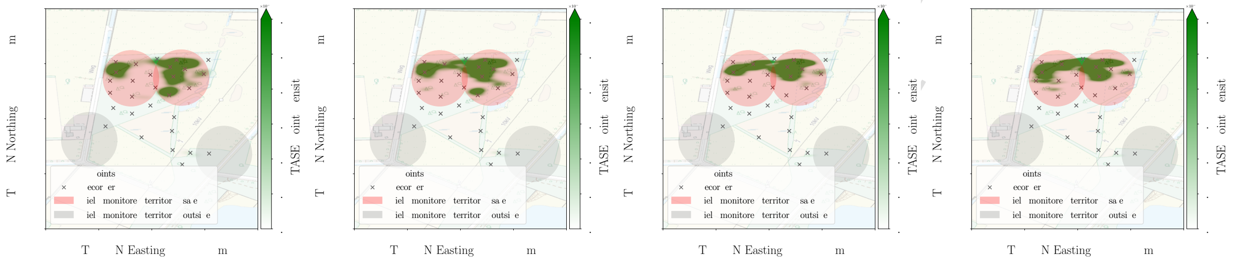
6.2. Influence of threshold$_R$ *in KDE Interpretation*

The following sections examine the relationship between KDE interpretation and the impact of the threshold *threshold*$_R$. Starting with the Eurasian *Sylvia atricapilla* in Fig. 7, we initially observe four distinct high-density clusters, separated by areas of lower density. These four clusters align closely with expert assessments. However, when we vary the threshold for TS roots *threshold*$_R$, as shown from Figs. 7a to 7d, the high-density area changes in form slightly. Note that with increasing *threshold*$_R$, the number of points declines from 6835 to only 1287. This reduction results from the interaction between the classifier's detection ability and TASE's assumption that any classification above *threshold*$_R$ is correct.

It is important to note that TASE detects individuals only when vocalizing. Since singing typically occurs at specific, localized spots rather than continuously across a territory, the resulting data may not form a cohesive area. Temporal or spatial gaps between vocalizations and a lack of observed movement can cause what would otherwise appear as a contiguous territory to break into several isolated "hotspots" of high-density activity. Examples of these separated hotspots are highlighted with orange arrows in Fig. 7. In Sec. 7.1, we further discuss how our methodology may also contribute to forming such fragmented areas.



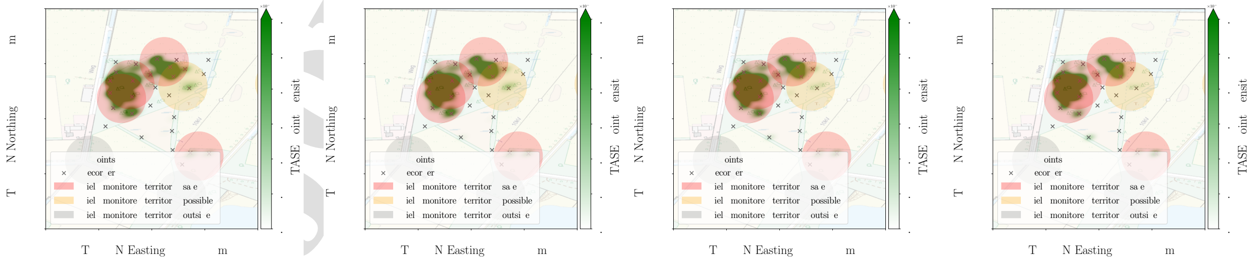
(a) 5:15:00 to 5:45:00 (b) 5:30:00 to 6:00:00 (c) 5:45:00 to 6:15:00 (d) 6:00:00 to 6:30:00



(e) 6:15:00 to 6:45:00 (f) 6:30:00 to 7:00:00 (g) 6:45:00 to 7:15:00 (h) 7:00:00 to 7:30:00

Figure 9: Kernel density estimate (green) of spatio-temporal detections for the Eurasian Wren *Troglodytes troglodytes* after TASE processing, with the expert's territorial assessment shown by the colored circles.

Independent of the $threshold_R$, when territories overlap in space or time, high-density areas can become large, as shown for the Eurasian Wren *Troglodytes troglodytes* in Fig. 8. Experts identify two distinct territories based on their knowledge of the species, but at first glance, the density map falsely suggests a single, larger territory. These distortions become visible by analyzing shorter time intervals, as illustrated in Fig. 9. For the Eurasian Wren, 30-minute segments with 15-minute overlaps reveal multiple high-density clusters, confirming that singing points are spaced apart rather than forming a single extensive area. Over time, the distance between these clusters decreases, reflecting a characteristic behavior known as counter-singing, where individuals advertise their territory boundaries toward neighboring birds. As a result, the gaps between territories gradually close, making it increasingly difficult to visually distinguish individual territories over long periods and highlighting the crucial role of expert knowledge.

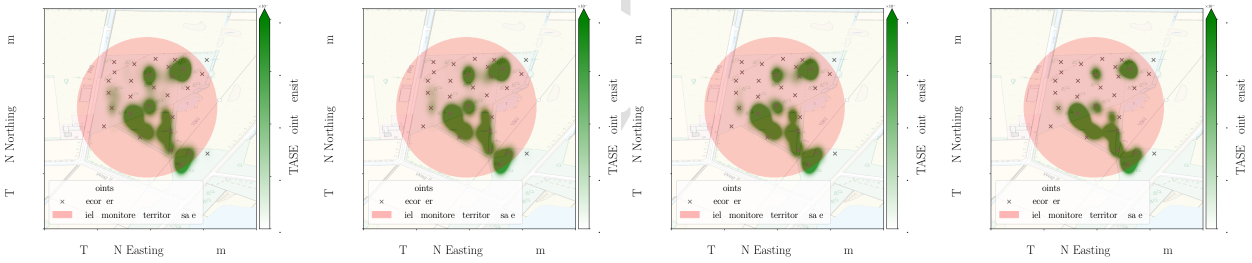


(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure 10: Kernel density estimate (green) of spatio-temporal detections for European Robin *Erithacus rubecula* after TASE, with the expert's territorial assessment shown by the colored circles.

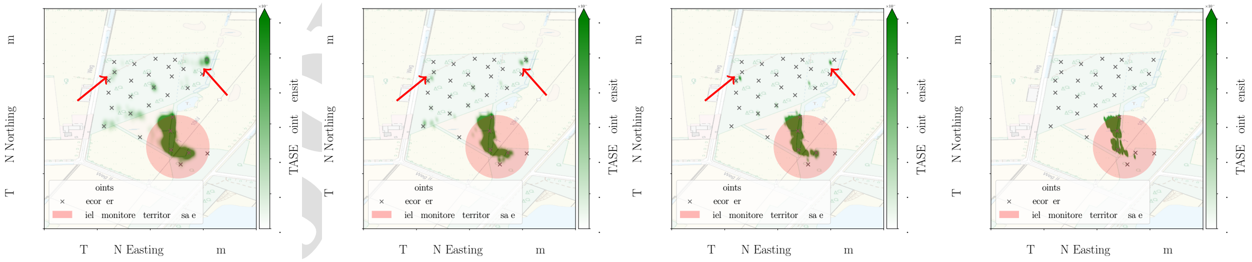
Reducing the time spans when visually interpreting high-density areas also has limitations, particularly when territories are very close together or overlap extensively, as seen with the European Robin *Erithacus rubecula* in Fig. 10. The three western territories merge into a single region without clear low-density borders, aligning with the expert-identified territories (red circles) but highlighting that TASE-based density estimation struggles to separate neighboring clusters. Dense territories are also challenging to assess in traditional field monitoring and are a known source of error [5], likely resolvable only through IAM. Our analysis indicates that using intervals shorter than 15 minutes can partially alleviate the issue. However, it also reveals that the challenges in assessing spatially close territories remain, underscoring the need for continued research in this area.

We identified two additional behavioral factors affecting the interpretation of our results. The first involves species with large territories. 1) Although expert field mapping confirmed the presence of a single territorial Song Thrush *Turdus philomelos*, its powerful, far-carrying song was recorded by multiple devices across the deployment site. As shown in Fig. 11, this resulted in multiple high-density clusters centered within the deployment area, making it difficult to distinguish clear territorial boundaries. 2) The second factor, which we have identified involves individuals who are either non-territorial or temporarily leave their established territories, thereby influencing the data interpretation. In Fig. 12, two verified high-density cluster (highlighted in red) appears simultaneously with vocalizations from another individual of the same species to the south, suggesting the involvement of multiple territories. Expert ground-truth data confirms only one territory within the deployment area, with two additional territories located approximately 300 meters beyond its boundaries. This suggests that territorial birds outside the study area may occasionally enter its perimeter or that non-territorial individuals may pass through. Such behavior is well-documented in ornithological studies and represents a recognized source of methodological error in field monitoring (e.g., [5]). These observations show that TASE can capture real behavioral dynamics, such as temporary territorial incursions and broader movements. It is clear that a more focused approach to spatial data analysis, particularly in terms of timing, could significantly enhance the interpretation of territories.



(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

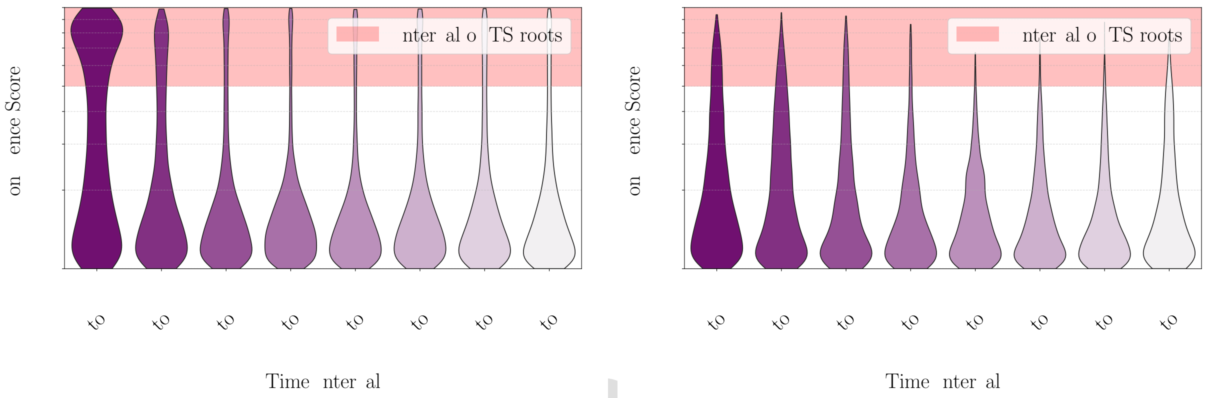
Figure 11: Kernel density estimate (green) of spatio-temporal detections for Song Thrush *Turdus philomelos* after TASE, with the expert's territorial assessment shown by the colored circles.



(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure 12: Kernel density estimate (green) of spatio-temporal detections for Tree Pipit *Anthus trivialis* after TASE, with the expert's territorial assessment shown by the colored circles.

The preceding evaluation demonstrates that TASE's results closely align with expert assessments of territorial boundaries. Similar findings for other species are presented in the appendix. Summarizing, territorial recognizability is affected by the interaction between the classifier performance, the *threshold*$_R$, and species behavior. (1) Lowering *threshold*$_R$ increases the inclusion of classifier errors in TASE, which could obscure territorial boundaries. Still, our evaluation shows that the overall pattern remains consistent. Selecting an appropriate *threshold*$_R$ depends on how well the classifier works for the species, the region, and the acoustic environment and should make territorial patterns detectable. (2) Territorial patterns can be observed, but separating them is challenging, especially when conspecifics are nearby. Shorter time-scale analyses can help by revealing local dynamics often masked over more extended periods. As already mentioned, this difficulty is not unique to TASE — field surveys encounter similar challenges ([49, pp. 52-53] [5]). Undoubtedly, expert knowledge of the target species' territorial behavior is important and essential to accurately interpret the KDE and minimize erroneous estimates of their number.



(a) Common Redstart (b) Song Thrush

Figure 13: Impact of overlapping vocalizations during dawn chorus, exemplary for two species

7. Discussion and Future Research

Our evaluation demonstrates that TASE, when applied to bird data collected by an ASN, can identify territorial patterns for multiple species coexisting in the same area. By visualizing these patterns, TASE enables experts to analyze the spatial distribution of high-density areas and accurately estimate the number of territories. This capability also unlocks new possibilities for monitoring species abundance, serving as a novel, powerful tool for advancing ecological research and conservation efforts.

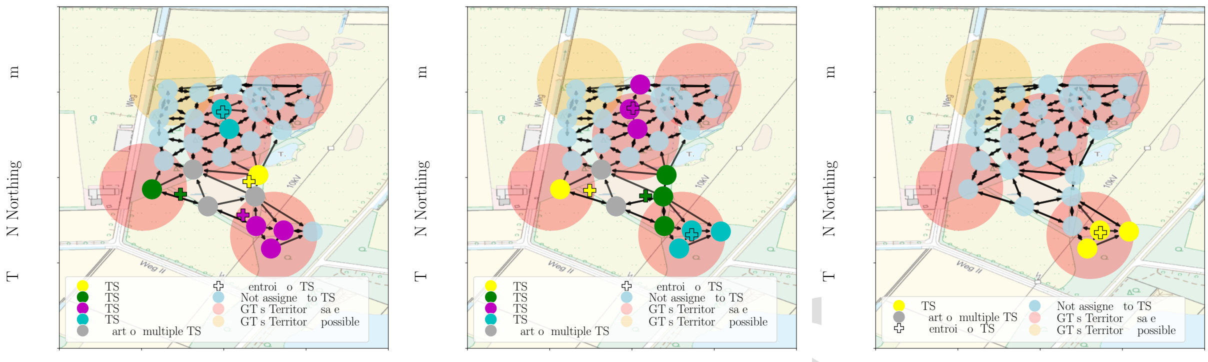
7.1. Limitations of TASE

7.1.1. Impact of Species Classifier

Our work presents a novel approach that leverages spatially distributed confidence values from the AI classifier BirdNet for species identification. Although BirdNet is widely used and relatively reliable, it struggles in complex environments like the dawn chorus [27]. Confidence values decrease due to overlapping sounds from multiple species sharing similar acoustic frequencies, which obscure each other's features.

Figure 13 shows the confidence scores per node for Common Redstart and Song Thrush detections in 15-minute bins. Values above 0.5 — seeds for territorial subgraphs — drop sharply between 04:00 and 04:30, when the dawn chorus intensifies. Before 04:30, high-confidence detections are far more common. Overlapping songs reduce confidence scores once additional species join in. This pattern applies to other species active during low-interference periods such as early morning and late evening. For TASE, due to its fixed *threshold*$_R$, this variation in confidences affects both the number of detection points and biases towards specific times — high thresholds favor low-interference periods.

A potential solution is to implement an adaptive $threshold_R$ that adjusts based on interference conditions, considering the confidence values of both target and other species to provide a more flexible threshold mechanism. Another option would be to ground-truth the acoustic data directly, enabling the calculation of precision and recall metrics that could support fine-tuned, species-specific thresholds tailored to the local acoustic environment. It is important to note that these metrics may also vary over time due to changing outdoor acoustic conditions and acoustic behavior.



(a) 04:00:5 to 04:00:8

(b) 04:00:41 to 04:00:43

(c) 04:37:48 to 04:37:51

Figure 14: Impact of variation in classifier's confidence over time.

7.1.2. Impact of TASE Parameters

Variations in the classifier's confidence scores can influence how well TASE generates spatio-temporal point clouds. To explore this effect, we examine three representative TASE runs in detail.

Local noise and interference can degrade classifier accuracy and can lead to overcounting. For example, the gray node between TS 1 and TS 3 in Fig. 14a has a confidence value of 0.582, causing it to be assigned to both territories and falsely splitting what should be a single TS into two. Additionally, we set a maximum distance from a TS's root node to its border nodes. However, during low-interference periods, high confidence levels can extend detections beyond this limit, resulting in duplicate TS, as illustrated in Fig.14b, where TS 2 (green) and TS 4 (cyan) represent the same individual. Interference also significantly reduces the detectable range of vocalizations, lowering classification confidence. As shown in Fig.14c, neighboring nodes recorded the Common Redstart with a maximum confidence of only 0.16, compared to much higher values under low-interference conditions (Figs.14a and 14b).

While these erroneous TS are an issue, our prior evaluation shows that they have minimal overall impact for two reasons: 1. We concentrate on the dawn chorus (4 to 10 a.m.), where low interference affects only about 45 minutes of the six-hour deployment, resulting in few induced errors. 2. The territorial nature of our species confines them to specific areas, minimizing error impact as long as territories are sufficiently spaced apart. Nevertheless, reducing classification variability can enhance our algorithm's resilience. A promising approach is to incorporate contextual information from species detection time series in the future, as proposed in [45].

7.1.3. Integrate precise Localization

In step 4 of TASE, we derive representations for the territorial subgraphs corresponding to individual vocalizing animals. A key advancement at this stage is the precise localization of these sources. Existing methods perform poorly in complex soundscapes like the dawn chorus, where overlapping vocalizations aggravate localization and hinder full automation[41]. However, prior work by [10] provides a promising approach for overcoming these challenges.

7.1.4. Automated Evaluation of spatio-temporal Point Cloud

Past evaluations have revealed that high-density areas, while aligning well with the territorial approach, also present significant limitations. These areas can evolve and change over time, with distinct territories merging into larger cohesive areas, posing a challenge in visual interpretation. It is crucial to recognize that territorial behavior is often more dynamic than the fixed-border assumption underlying this work. This shift in perspective is necessary for a more comprehensive understanding of the subject (see Sec. 7.2). Future studies should adopt more sophisticated clustering methods that account for spatial and temporal scales and are capable of handling occasional errors in the data or migrating individuals. Incorporating expert knowledge into the point cloud clustering process without focusing only on the point density is the long-term direction for automated evaluation. However, translating expert insights—such as species behavior or habitat use—into algorithmic rules remains challenging. Strategies like semi-supervised learning or rule-based constraints may help, though they could increase computational complexity.

7.2. Territoriality - Limitations and Future Research

While territorial spatial patterns generally align with expert assessments, TASE struggles when territories overlap extensively, when a single territory is unusually large, or when territorial individuals migrate through the array altogether — difficulties that also arise in traditional field monitoring. Our findings suggest that spatial patterns of individuals might exceed the idea of fixed territorial boundaries., which is already well-documented in ornithology (e.g., [5], [49, pp.52]), where it is known that large territories, densely populated habitats, and individuals temporarily leaving their core territories can complicate assessment of territories. By comparing these field-based methodological challenges with TASE's performance, we find that both methods share similar limitations. However, we also find that many field monitoring limitations are caused by data scarcity. We expect TASE to overcome those since recorders remain in the field continuously, collecting data when species are undisturbed. But this advantage brings a new challenge: developing advanced analytical methods to untangle complex territorial dynamics. We see two primary research directions to address them: 1. Extend the fixed territory concept by a soft boundary concept: Statistical models could define territorial subgraphs by embracing uncertainty and allowing for dynamic, overlapping boundaries. 2. Replace the fixed temporal resolution with a dynamic one: Using fixed time intervals may not capture species-specific temporal dynamics. Adaptive temporal resolution could sharpen territorial estimates, especially for species with fluctuating activity. Even so, expert insight remains crucial for interpreting the dynamics and extending TASE to longer deployments in diverse acoustic settings.

7.3. Future Viability of Next-Generation Species Classifiers

New, more advanced species identification models, such as those introduced by [24], are emerging and may eventually surpass BirdNET's current performance, including improved robustness to overlapping sounds. As these classifiers become more resilient to interference, TASE will benefit accordingly. However, TASE relies on a fundamental assumption: there must be a correlation between classification confidence and distance. Before adopting any new classifier, verifying that it meets this criterion is essential. Datasets like [50] or Open-Source acoustic simulators such as [8] can assist in conducting these preliminary evaluations to ensure that future classifiers fulfill that correlation.

7.4. Time- and Space Scalability

The deployment in this paper is limited in time and space. To assess TASE performance, longer and larger deployments should be conducted on various species. Of course, the computational demand grows with number of sensor nodes (spatial coverage) and input temporal resolution. TASE must be run on much broader data sets to capture richer, species-specific territorial dynamics that will emerge over longer time spans and larger areas. Reducing computational time by parallelizing the algorithm is an objective for future work.

In addition to the TASE application, ground-truthing also becomes an issue. Manual validation becomes impractical when scaling up to larger or more complex environments, such as a rainfores. Given the limitations of TASE application and the impracticality of manual validation in such environments, there is an need to explore alternative ground-truthing methods. Future research might focus on methods such as bird banding, GPS tracking, or semi-automated IAM techniques for broader applications.

Our approach relies on densely placed recorders, which raises several practical questions. Therefore, future work should prioritize developing clear deployment guidelines for ornithologists because current knowledge is based on practical experience. Key priorities include: (i) optimizing sampling periods to balance battery life, storage, and detection rates, (ii) standardizing microphone gain and configuration settings to ensure consistent recording quality, (iii) establishing practical field deployment strategies, including recommendations for sensor spacing, mounting, and orientation.

7.5. Transferability to other Species and Regions

Although we applied TASE primarily to birds, the underlying approach can be adapted to other territorial, sound-producing species, such as cicadas [52]. The key prerequisites remain the same: (i) the species is territorial and emits identifiable vocalizations, (ii) a reliable classifier exists, and (iii) the requirements in Sec. 4.1 are satisfied. Detailed knowledge of a species' call structure and territorial behavior remains critical because it dictates recorder spacing and the spatial parameters of the detection graph. Once these inputs are in place, the workflow transfers with minimal extra effort.

Scaling TASE to large, species-rich soundscapes, such as tropical forests, poses additional challenges that TASE likely fails to handle. For example, the "one bird per cluster" assumption, which is based on the idea that each bird's vocalization can be distinctly identified and separated from others, may fail when several individuals sing in close proximity. Possible solutions might involve denser node deployments with limited recording ranges or directional microphones, which eventually require modifications when building the graph.

7.6. Opening up new research Opportunities

TASE advances bioacoustic monitoring on two critical fronts. *First*, accurately estimating territories through bioacoustic methods offers significant ecological and conservation benefits. By delineating the spatial extent of different individuals or populations, researchers gain insights into species distribution, resource use, and habitat preferences. Such knowledge can inform land-use decisions, guide habitat restoration efforts, and improve species management strategies. When combined with long-term acoustic monitoring, territorial estimates can help to track population dynamics, detect changes in species abundance or distribution over time, and identify critical areas for conservation. *Second*, TASE collects reliable abundance data in places where traditional field surveys cannot operate, e.g., remote alpine valleys or countries lacking extensive bird-watching expertise. Because many species migrate across national borders, closing these data gaps is essential for coordinated, multi-country protection efforts.

8. Conclusion

This work offers a promising solution for estimating species abundance from ASNs using an AI-based species classifier. Instead of identifying individuals by their unique features, we leverage their territoriality to estimate the number of territories as a proxy for species abundance. Our approach combines a species classifier with a network of acoustic recorders deployed across a cohesive area to reveal spatial patterns corresponding to territories. We tested our method on territorial bird species in a qualitative proof-of-concept evaluation. Our results were compared with expert field assessments, providing a robust validation of our approach. Our findings closely matched expert perceptions of territorial boundaries despite errors arising from complex avian behaviors and classifier limitations. However, much refinement remains, including assessing TASE in large-scale and long-term deployments and introducing a concept of soft-territorial boundaries to develop automated spatial-temporal clustering instead of spatial high-density interpretation. In future research, TASE has the potential to become a valuable tool for acquiring abundance data in areas where field monitoring is traditionally challenging.

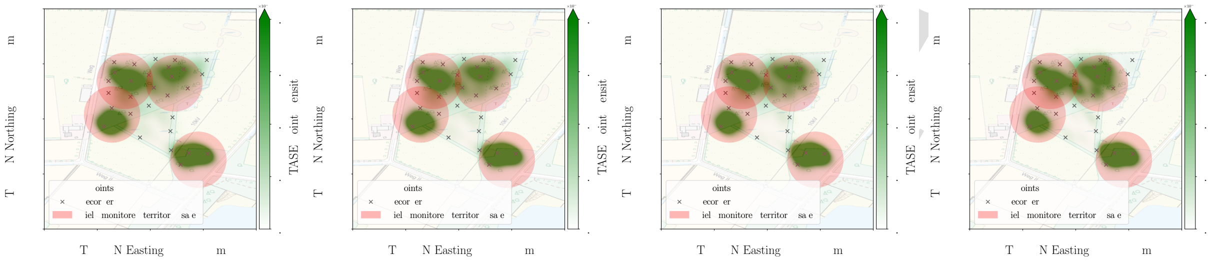
Data Availability

We provide our dataset at the following link: `https://doi.org/10.26249/FK2/VORDOF` Companion tools, including our algorithm and plotting scripts, are available at: `https://github.com/sys-uos/TASE`

Declaration of generative AI and AI-assisted technologies in the writing process

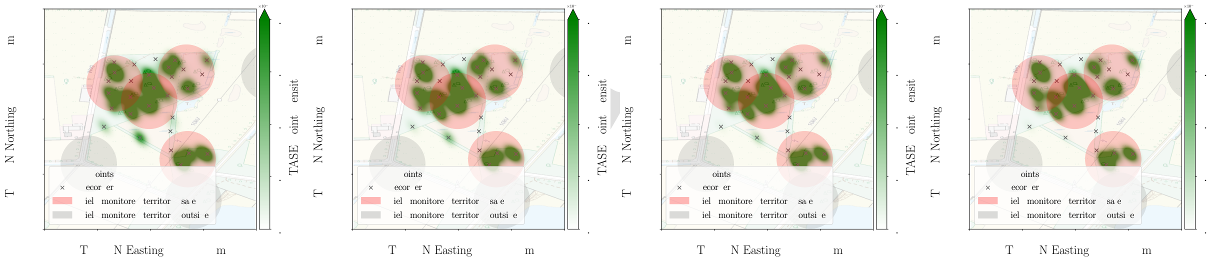
The authors used ChatGPT and Grammarly to enhance the manuscript's readability and language, thoroughly reviewing and revising the content, and assume full responsibility for the final published article.

A. Appendices



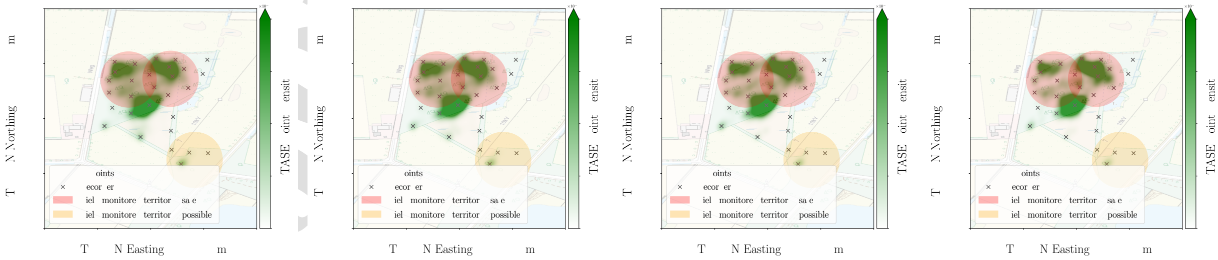
(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure A.1: Kernel density estimate (green) of spatio-temporal detections for Short-toed Treecreeper *Certhia brachydactyla* after TASE, with the expert's territorial assessment shown by the colored circles.



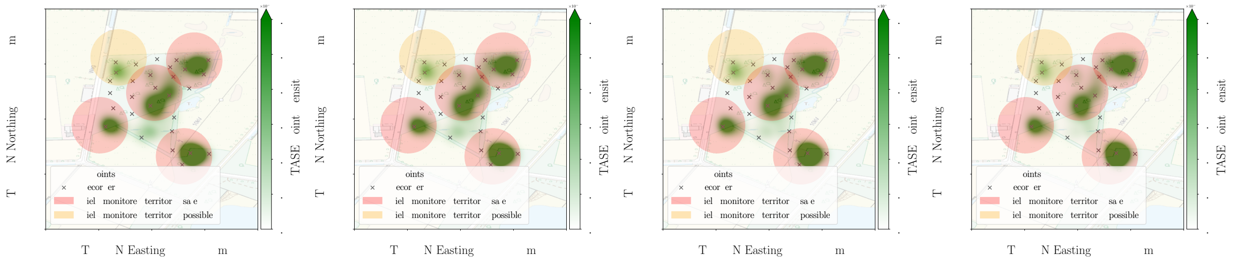
(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure A.2: Kernel density estimate (green) of spatio-temporal detections for Eurasian Chaffinch *Fringilla coelebs* after TASE, with the expert's territorial assessment shown by the colored circles.



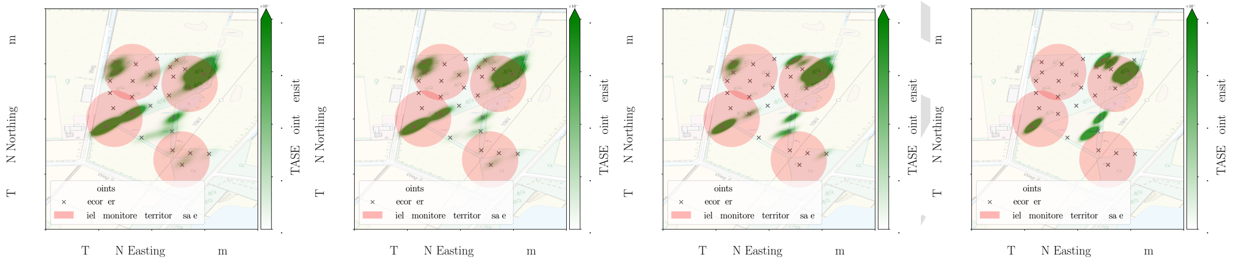
(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure A.3: Kernel density estimate (green) of spatio-temporal detections for Spotted Flycatcher *Muscicapa striata* after TASE, with the expert's territorial assessment shown by the colored circles.



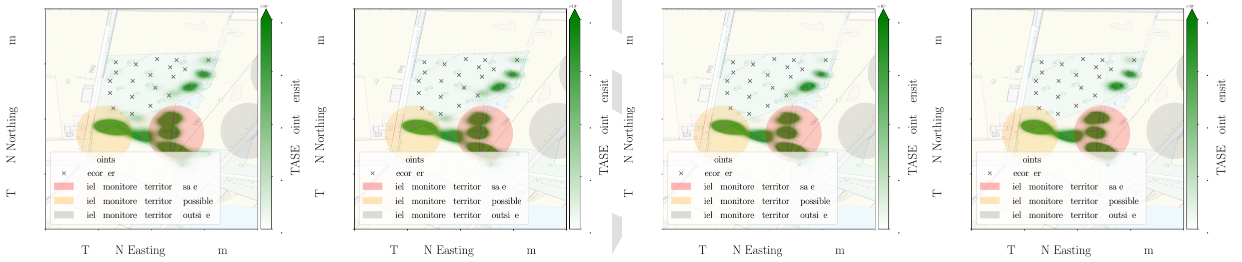
(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure A.4: Kernel density estimate (green) of spatio-temporal detections for Common Redstart *Phoenicurus phoenicurus* after TASE, with the expert's territorial assessment shown by the colored circles.



(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure A.5: Kernel density estimate (green) of spatio-temporal detections for Common Chiffchaff *Phylloscopus collybita* after TASE, with the expert's territorial assessment shown by the colored circles.



(a) $threshold_R$: 0.5 (b) $threshold_R$: 0.6 (c) $threshold_R$: 0.7 (d) $threshold_R$: 0.8

Figure A.6: Kernel density estimate (green) of spatio-temporal detections for Garden warbler *Sylvia borin* after TASE, with the expert's territorial assessment shown by the colored circles.

References

[1] Abrahams, C., and Denny, M. J. H. A first test of unattended, acoustic recorders for monitoring Capercaillie Tetrao urogallus lekking activity. *Bird Study 65*, 2 (2018), 197–207.

[2] Adi, K., Johnson, M. T., and Osiejuk, T. S. Acoustic censusing using automatic vocalization classification and identity recognition. *The Journal of the Acoustical Society of America 127*, 2 (2010), 874–883.

[3] Amrhein, V., Kunc, H., and Naguib, M. Seasonal patterns of singing activity vary with time of day in the nightingale (luscinia megarhynchos). *The Auk 121* (01 2004), 110–117.

[4] Bauer, H.-G., Bezzel, E., and Fiedler, W. *Das Kompendium der Vögel Mitteleuropas: Alles über Biologie, Gefährdung und Schutz.* AULA-Verlag, Wiebelsheim, Hunsrück, 2011.

[5] Berthold, P. Methoden der Bestandserfassung in der Ornithologie: Übersicht und kritische Betrachtung. *Journal für Ornithologie 117*, 1 (1976), 1–69.

[6] Bezzel, E., and Prinzinger, R. *Ornithologie.* UTB Große Reihe. Ulmer Verlag, Stuttgart, 1990.

[7] Brackenbury, J. H. Power Capabilities of the Avian Sound-producing System. *Journal of Experimental Biology 78*, 1 (1979), 163–166.

[8] Brüggemann, L., Schütz, B., and Aschenbruck, N. TAWNS – A Terrestrial Acoustic and Wireless Network Simulation Framework. In *Performance Evaluation Methodologies and Tools* (2023), Springer Nature Switzerland, pp. 292–306.

[9] Brumm, H. Song amplitude and body size in birds. *Behavioral Ecology and Sociobiology 63*, 8 (2009), 1157–1165.

[10] Brüggemann, L., Dyczka, M., Otten, D., and Aschenbruck, N. A classifier-deduced signal extraction approach for time difference estimation in acoustic sensor networks. *Ecological Informatics 87* (2025), 103032.

[11] Budka, M., Jobda, M., Szałański, P., and Piórkowski, H. Acoustic approach as an alternative to human-based survey in bird biodiversity monitoring in agricultural meadows. *PLOS ONE 17*, 4 (2022).

[12] Cheng, J., Xie, B., Lin, C., and Ji, L. A comparative study in birds: Call-type-independent species and individual recognition using four machine-learning methods and two acoustic features. *Bioacoustics 21*, 2 (2012), 157–171.

[13] Chhaya, V., Lahiri, S., Jagan, M. A., Mohan, R., Pathaw, N. A., and Krishnan, A. Community Bioacoustics: Studying Acoustic Community Structure for Ecological and Conservation Insights. *Frontiers in Ecology and Evolution 9* (2021).

[14] Chronister, L. M., Rhinehart, T. A., and Kitzes, J. When birds sing at the same pitch, they avoid singing at the same time. *Ibis 165*, 3 (2023), 1047–1053.

[15] Darras, K., Furnas, B., Fitriawan, I., Mulyani, Y., and Tscharntke, T. Estimating bird detection distances in sound recordings for standardizing detection ranges and distance sampling. *Methods in Ecology and Evolution 9*, 9 (2018), 1928–1938.

[16] Diego, A., Gómez Catasús, J., Pérez Granados, C., Bustillo de la Rosa, D., and Traba, J. Conspecific density and habitat quality drive the defence and vocal behaviour of a territorial passerine. *Ibis 166*, 3 (2023), 826–843.

[17] Foote, J. R., Palazzi, E., and Mennill, D. J. Songs of the Eastern Phoebe, a suboscine songbird, are individually distinctive but do not vary geographically. *Bioacoustics 22*, 2 (2013), 137–151.

[18] Fox, Elizabeth., Roberts, D., and Bennamoun, M. Call-independent individual identification in birds. *Bioacoustics 18*, 1 (2008), 51–67.

[19] Fraixedas, S., Lindén, A., Piha, M., Cabeza, M., Gregory, R., and Lehikoinen, A. A state-of-the-art review on birds as indicators of biodiversity: Advances, challenges, and future directions. *Ecological Indicators 118* (2020), 106728.

[20] Funosas, D., Barbaro, L., Schillé, L., Elger, A., Castagneyrol, B., and Cauchoix, M. Assessing the potential of birdnet to infer european bird communities from large-scale ecoacoustic data. *Ecological Indicators 164* (2024), 112146.

[21] Ghani, B., Denton, T., Kahl, S., and Klinck, H. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports 13*, 1 (2023), 22876.

[22] Gros-Desormeaux, H., Hunel, P., Vidot, N., and Stattner, E. Acoustic counting algorithms for wireless sensor networks. In *Proceedings of the 6th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks* (2009), PE-WASUN '09, Association for Computing Machinery, p. 79–84.

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

[24] Höchst, J., Bellafkir, H., Lampe, P., Vogelbacher, M., Mühling, M., Schneider, D., Lindner, K., Rösner, S., Schabo, D. G., Farwig, N., and Freisleben, B. Bird@Edge: Bird Species Recognition at the Edge. In *Networked Systems* (2022), M.-A. Koulali and M. Mezini, Eds., Springer International Publishing, pp. 69–86.

[25] Howard, E. *Territory in Bird Life*, new edition ed. Wm. Collins, Sons, and Co., Ltd., London and Glasgow, 1948.

[26] International Committee of Bio-acoustics. *Acoustic Behaviour of Animals.* Elsevier, 1963.

[27] Kahl, S., Wood, C. M., Eibl, M., and Klinck, H. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics 61* (2021).

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

[29] Lauha, P., Somervuo, P., Lehikoinen, P., Geres, L., Richter, T., Seibold, S., and Ovaskainen, O. Domain-specific neural networks improve automated bird sound recognition already with small amount of local data. *Methods in Ecology and Evolution 13*, 12 (2022), 2799–2810.

[30] Lee, D. T., and Schachter, B. J. Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences 9*, 3 (1980), 219–242.

[31] Maher, C. R. Territoriality. In *Encyclopedia of Animal Cognition and Behavior*, J. Vonk and T. K. Shackelford, Eds. Springer International Publishing, 2022, pp. 6910–6918.

[32] Maher, C. R., and Lott, D. F. A Review of Ecological Determinants of Territoriality within Vertebrate Species. *The American Midland Naturalist 143*, 1 (2000), 1–29.

[33] Navine, A. K., Camp, R. J., Weldy, M. J., Denton, T., and Hart, P. J. Counting the chorus: A bioacoustic indicator of population density. *Ecological Indicators 169* (2024), 112930.

[34] Navine, A. K., Denton, T., Weldy, M. J., and Hart, P. J. All thresholds barred: direct estimation of call density in bioacoustic data. *Frontiers in Bird Science 3* (2024).

[35] Oñate-Casado, J., Porteš, M., Beran, V., Petrusek, A., and Petrusková, T. Guess who? Evaluating individual acoustic monitoring for males and females of the Tawny Pipit, a migratory passerine bird with a simple song. *Journal of Ornithology 164* (2023).

[36] O'Donald, P. Sexual Selection and the Evolution of Territoriality in Birds. In *Measuring Selection in Natural Populations* (1977), F. B. Christiansen and T. M. Fenchel, Eds., Springer, pp. 113–129.

[37] Pérez-Granados, C. A First Assessment of Birdnet Performance at Varying Distances: A Playback Experiment. *Ardeola 70*, 2 (2023), 257–269.

[38] Pérez-Granados, C., and Traba, J. Estimating bird density using passive acoustic monitoring: A review of methods and suggestions for further research. *Ibis 163*, 3 (2021), 765–783.

[39] Potts, J. R., and Lewis, M. A. How do animal territories form and change? Lessons from 20 years of mechanistic modelling. *Proceedings of the Royal Society B: Biological Sciences 281*, 1784 (2014), 20140231.

[40] Ptacek, L., Machlica, L., Linhart, P., Jaska, P., and Muller, L. Automatic recognition of bird individuals on an open set using as-is recordings. *Bioacoustics 25*, 1 (2016), 55–73.

[41] Rhinehart, T. A., Chronister, L. M., Devlin, T., and Kitzes, J. Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecology and Evolution 10*, 13 (2020).

[42] Ross, S. R. P.-J., O'Connell, D. P., Deichmann, J. L., Desjonquères, C., Gasc, A., Phillips, J. N., Sethi, S. S., Wood, C. M., and Burivalova, Z. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Functional Ecology 37*, 4 (2023), 959–975.

[43] Schlüpmann, M., and Kupfer, A. Methoden der amphibienerfassung – eine Übersicht. *Zeitschrift für Feldherpetologie, Supplement 15* (2009), 7–84.

[44] Serrurier, A., Zdroik, P., Isler, R., Kornienko, T., Peris-Morente, E., Sattler, T., and Pradervand, J.-N. Mountain is calling – decrypting the vocal phenology of an alpine bird species using passive acoustic monitoring. *Ibis 166*, 4 (2024), 1338–1353.

[45] Singer, D., Hagge, J., Kamp, J., Hondong, H., and Schuldt, A. Aggregated time-series features boost species-specific differentiation of true and false positives in passive acoustic monitoring of bird assemblages. *Remote Sensing in Ecology and Conservation 10*, 4 (2024), 517–530.

[46] Singer, D., Kamp, J., Hondong, H., Schuldt, A., and Hagge, J. Diel and seasonal vocal activity patterns revealed by passive acoustic monitoring suggest expert recommendations for breeding bird surveys need adjustment, 2024. Preprint.

[47] Stattner, E., Hunel, P., Vidot, N., and Collard, M. Acoustic scheme to count bird songs with wireless sensor networks. In *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks* (2011), pp. 1–3.

[48] Stowell, D., Petrusková, T., Šálek, M., and Linhart, P. Automatic acoustic identification of individuals in multiple species: Improving identification across recording conditions. *Journal of The Royal Society Interface 16*, 153 (2019), 20180940.

[49] Südbeck, P., Andretzke, H., Gedeon, K., Schikore, T., Schröder, K., Fischer, S., and Sudfeldt, C. *Methodenstandards Zur Erfassung Der Brutvögel Deutschlands*. Max-Planck-Institut für Ornithologie. Vogelwarte Radolfzell, 2005/2012.

[50] Winiarska, D., Szymański, P., and Osiejuk, T. S. Detection ranges of forest bird vocalisations: Guidelines for passive acoustic monitoring. *Scientific Reports 14*, 1 (2024), 894.

[51] Wood, C. M., and Kahl, S. Guidelines for appropriate use of BirdNET scores and other detector outputs. *Journal of Ornithology 165*, 3 (2024), 777–782.

[52] Yohanes, G., Prabowo, A. S., and Kurniadi, F. I. Cicada species recognition based on acoustic signals using dynamic time warping-graph based GraphMix, graph convolution network. *Procedia Computer Science 245* (2024), 508–517.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: