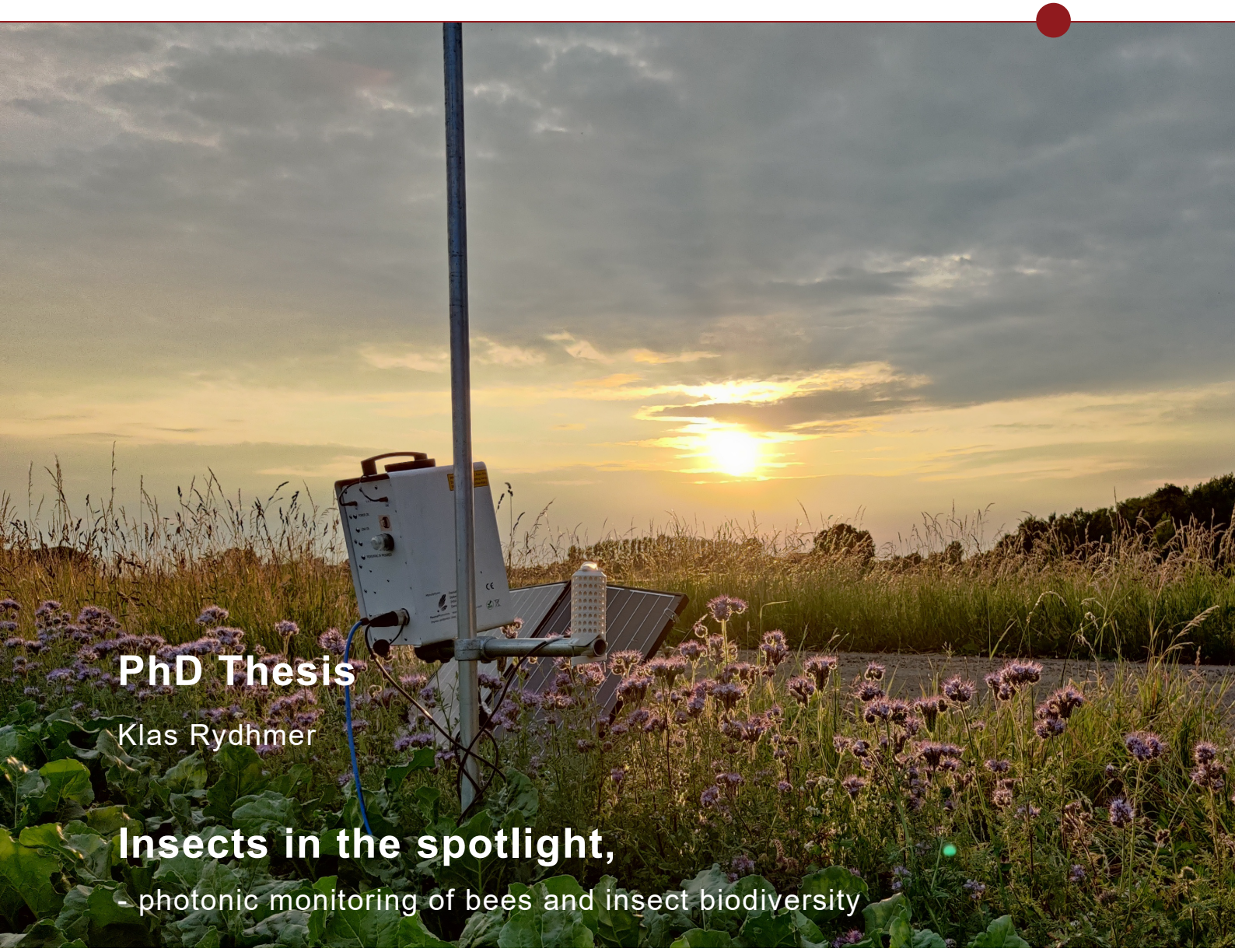
UNIVERSITY OF COPENHAGEN
DEPARTMENT OF GEOSCIENCES AND NATURAL RESOURCE MANAGEMENT



PhD Thesis

Klas Rydhmer

Insects in the spotlight,

- photonic monitoring of bees and insect biodiversity

Supervisor: Inger Kappel Schmidt

Submitted on: 30/6 2023

This thesis has been submitted to the PhD school of The Faculty of Science, University of Copenhagen

Name of department:	Department of Geosciences and Natural Resource Management
Author(s):	Klas Rydhmer
Title and subtitle:	Insects in the spotlight -photonic monitoring of bees and insect diversity
Topic description:	Automated insect monitoring
Supervisor:	Inger Kappel Schmidt
Co-supervisors:	Mikkel Brydegaard, Henrik Smith, Carsten Kirkeby
Company supervisors:	Mads Fogtmann, (Thomas Nikolajsen, Jesper Lemmich, Flemming Rasmussen)
Submitted on:	June 30th 2023.

Preface

After working as a research engineer and data scientist at FaunaPhotonics for a couple of years, I was offered the possibility to change my employment into an industrial PhD position. At that point, we had focused on aquatic salmon lice, detection of birds around windmills and insect pest monitoring in oilseed rape. The development of the first “commercially ready” insect sensor was just started, and the company had grown from five persons when I joined into more than twenty people. Collecting labelled data on new pest species and expanding to more crops were supposed to be routine work in coming years.

I was offered a chance to focus on a new area, working closer to researchers rather than agrochemical business partners and we initialized biodiversity monitoring as a side track to complement the more commercially important pest monitoring. As we wrote the application, OECD published their report “*Biodiversity: Finance and the Economic and Business Case for Action”* and suddenly biodiversity monitoring was in every headline.

The side track offered a bumpy ride. As I am about to submit the thesis, biodiversity monitoring is proposed to be mandatory by EU law. Rather than being a slow and steady academic side project, it is now the key focus of FaunaPhotonics. Only one of the first four colleagues remain and I am the oldest employee in the company. The project faced Covid, three company supervisor replacements, the birth of my first child and most recently war in Europe. It has been three eventful, awful and wonderful years.

/Klas, June 2023

Table of Contents

Abstract

This thesis is the result of an industrial PhD project in cooperation between FaunaPhotonics A/S and the Department of Geosciences and Natural Resource Management at Copenhagen University. It explores the viability of using photonic sensors for monitoring of bees and insect diversity.

Recently reported declines of insect abundance and diversity have illustrated the need for large scale and long running studies of insects. However, such studies are costly with conventional methods which are constrained in their temporal and spatial coverage. Thus, there is a need for new methods to complement the conventional monitoring approaches. Entomological lidar has become increasingly common in recent years and allows the recording of thousands of insect observations in minutes. Such instruments emit a beam of light and record the reflected light from insects passing through the beam. The data is recorded as a time signal where, amongst other features, the wingbeat frequency is seen as a modulation in the signal.

FaunaPhotonics have developed a short-range optical instrument with a similar measurement principle to entomological lidar. The "Volito" sensor automatically record and extract insect observations from the data and transmit them to a cloud platform via cellular network. It is eye-safe, weatherproof and capable of long unsupervised deployments in the field. While initially developed for pest monitoring in agricultural crops, the non-intrusive monitoring method makes it a suitable technology for general insect monitoring.

Monitoring biodiversity is a complex task that can involve identifying a large number of insects to sub-species level. While this seems unfeasible with optical instrumentation, this PhD project explores the possibility to develop a "simple" biodiversity indicator, analogous to species richness, or Simpson´s biodiversity index.

This thesis includes a spatial model for honeybee monitoring using conventional entomological lidar, showing good agreement with manual observations. It describes the development of the new "Volito" sensor and correlates measured insect abundance with yellow water traps. It explores a deep learning approach to feature extraction from recorded data, improving upon earlier models from literature. Finally, it includes two draft manuscripts covering the first attempts at correlating a sensor-based diversity metric with conventional monitoring methods in agricultural fields in Iowa, USA, as well as protected areas in southern Scandinavia.

Resumé

Denne afhandling er resultatet af et industrielt PhD projekt udført som et samarbejde mellem FaunaPhotonics A/S og Institut for Geovidenskab og Naturforvaltning på Københavns Universitet. I afhandlingen undersøges muligheden for brugen af optiske sensorer til monitorering af bier og insektdiversitet.

De seneste års rapportering af fald i antallet af insekter og diversitet har understreget nødvendigheden af længerevarende og omfattende insektstudier. Sådanne studier er dog omkostningstunge ved brug af konventionelle metoder, som desuden er begrænsede i deres tids- og arealmæssige omfang. Der er derfor behov for nye metoder til at komplimentere de konventionelle målemetoder. Entomologisk lidar er blevet stadig mere udbredt i de seneste år og muliggør måling af tusindvis af insekter på få minutter. Instrumenterne udsender en stråle af lys og måler det reflekterede lys fra insekter, der flyver gennem strålen. Data registreres som et tidssignal hvor forskellige karakteristika, heriblandt insektets vingeslagsfrekvens, kan findes som en variation i signalet.

FaunaPhotonics har udviklet et optisk instrument til måling over kortere afstande, som bygger på måleprincipper fra entomologisk lidar. Denne "Volito" sensor måler og isolerer automatisk insektobservationer fra data, og sender dem derefter til en platform i skyen via mobil netværk. Den er sikker for øjnene, vandtæt og kan klare langvarig udsætning i felten uden overvågning. Selvom den oprindeligt blev udviklet til overvågning af skadedyr i afgrøder, udgør denne målemetode en anvendelig teknologi til monitorering af insekter generelt uden at inteferere med insekterne.

Monitorering af biodiversitet er en kompleks udfordring, som kan indebære identifikation af et stort antal insekter helt ned til sub-artsniveau. Skønt dette umiddelbart kan forekomme udfordrende med optisk instrumentering, så afdækker dette PhD projekt muligheden for at udvikle en "simpel" biodiversitetsindikator, tilsvarende artsrigdom eller Simpson's biodiversitetsindex.

Denne afhandling indeholder en spatial model for monitorering af honningbier med konventionel entomologisk lidar, der viser en god overensstemmelse med manuelle observationer. Den beskriver udviklingen af den nye "Volito" sensor og korrelerer antallet af målte insekter med antallet af insekter fanget i gule fangbakker. Den undersøger en deep learning tilgang til at finde karakteristika fra det opsamlede data, forbedret fra tidligere modeller beskrevet i litteraturen. Til sidst indeholder den to udkast til manuskripter der dækker de første forsøg på at korrelere en

sensorbaseret diversitetsenhed med konventionelle overvågningsmetoder i dyrkede marker i Iowa, USA, samt natur områder i den sydlige del af Skandinavien.

General introduction

Insects play a crucial role in ecosystems as pollinators of wild plants and serve as food for larger animals such as birds (1). They also have a direct impact on human life by acting as pollinators, pests and pest control in agricultural systems and are common disease vectors (1–4). There is concern that a decline in insect abundance and diversity can cause cascade effects across the entire ecosystem at large scales (3). Recent reports of declining insect abundance and community health have generated a lot of attention and worries both the public as well as decision makers and academics (5).

The decline in insect abundance and diversity is mainly considered an effect of changes in agricultural practices and landscapes (6–8), primarily from increased use of insecticide, habitat losses from agricultural intensification and reduced grazing areas. However, there are few studies that have evaluated insect abundance across long temporal and large spatial scales since such studies are resource intensive with conventional methods (9).

Conventional insect diversity monitoring is done by trapping where insects are attracted by visual or olfactory cues or intercepted in flight or sweep netting (10). Collected specimen are stored in alcohol for later identification by microscope. As each trap has a different bias, a plethora of traps and monitoring methods are needed to get the unbiased insect diversity (10,11). An alternative, less intrusive method is to conduct the identification directly in flight in the field by Pollard walks or using a stationary observer (12). However, this limits the taxonomic accuracy and is only viable for larger species. Pest monitoring is, in general, easier than insect diversity monitoring as it is focused on a few key species which often can be counted directly in traps or on plants directly in the field (13).

As conventional methods either provide a very low temporal resolution, as with traps, or single “snapshots”, as with manual observation, they are difficult to apply at large scale and over long timeframes. While field visits are time consuming, the taxonomic identification of collected specimen is in general the most resource intensive part (14,15). Recent advances in DNA based methods could reduce these costs but the methods are not fully developed and are currently unable to provide species specific abundance information (15,16). In order to get accurate information on insect communities and study the effects of efforts to reverse the current biodiversity declines, there is a need for novel methods (15).

Automated insect monitoring

Automated insect monitoring methods can provide insect data without repeated human visits to the field sites (15). This could help researchers study the cause and effect of biodiversity promoting or pest supressing efforts as well as long-term trends, displacement effects and *diel* patterns, that are hard to capture with current methodology (17,18).

Automated insect traps for pest monitoring are commercially available today (19,20). Such traps typically use a pheromone lure to attract specific species or taxonomic groups and collected insects are identified by an optical system, such as a camera or photodiode sensor, in the trap (21,22). While the lure increases the catch and limits the range of species that enters the trap, it makes the method unsuitable for insect diversity monitoring.

Insect communities can also be monitored by remote sensing methods using optics, acoustics or radar. Camera systems can be used outside of traps but are limited by resolution and depth of focus (23,24). To allow identification to species level, high quality images are needed which either requires expensive components, or limited sampling volumes (25). Successful implementations typically focus the cameras on a small area, such as a flower, or an attractant such as a bright plastic piece, or an illuminated sheet (26,27). It is difficult to capture sufficiently high-quality images of free flying insects using machine vision systems, but some systems aimed at general insect monitoring are reported to be in development (28).

Acoustic methods have shown promising results, but are in general limited to audible insects, such as crickets, bees and mosquitoes (29–31). However, acoustic monitoring methods can also be deployed in solid mediums such as trees or grain silos (32–34). An advantage is that they can be used in citizen science, as any smartphone can serve as a recorder albeit with low range. As the acoustic methods primarily use the wingbeat frequency as an indicator, the specificity is limited as the wingbeat frequency of many insect species are overlapping (35–38).

Wingbeat frequencies can also be recorded by radar or lidar, as the projection, or optical cross section, of insect changes with the wing movement. Such methods can complement the wingbeat frequency with additional features, based on the optical properties of the insects (35,37–39). In general such instrumentation have large monitoring volumes and can record up to hundreds of thousands of insect observations in a single day (17). However, the cost and technical skill required to operate these instruments is currently high.

Radar and lidar entomology

Insects were observed as noise in radar data already in 1947 (40). At the time referred to as "angels", these observations were shortly confirmed to be reflections from flying insects (41). Dedicated entomological radar systems were developed in 1969 (42) and since then entomological radars have become common (43). Entomological radars typically emit radiation in the 2 – 110 GHZ range (corresponding to wavelengths of 10 cm down to 3 mm) as pulses and record the back-scattered signal from insects traversing the emitted beam. To avoid scatter from terrain and vegetation, they are typically aimed above the horizon and are able to record insects flying at kilometre ranges (43). By wobbling the beam direction, the flight direction and speed can be estimated and entomological radars are frequently used to monitor migratory insects such as moths (43,44). Recent advantages have also allowed existing networks of atmospheric doppler radars to be used for biodiversity monitoring of birds as well as insects (45,46).

Interference from terrain and vegetation can be reduced by using harmonic radar. By attaching harmonic reflectors to individual insects, flight and foraging behaviour can be recorded at local level (43,47,48). However, the technology can only be applied to insects sufficiently large to carry the reflector, such as bees. The beam divergence, and thus the ground scatter, can also be reduced by reducing the emitted wavelength (25). In comparison, entomological lidar operates on the same principles as entomological radar but rather than emitting radar waves, it uses laser beams, which can stay collimated at astronomical distances (49).

Development of the entomological lidar

Optical insect monitoring was demonstrated already in 1949 in Copenhagen, where a searchlight was used to monitor the nocturnal flight behaviour of moths (43,50) but entomological lidar was first demonstrated in 2005 to monitor bees trained to find land mines (51). By replacing the emitted radiation with a collimated pulsed laser beam, the instrument could operate near ground level and by panning the beam, provide a heat map of insect activity (and landmines).

Many current entomological lidar implementations use diode lasers and photodiode or CMOS detectors rather than solid state lasers and photo multiplier tubes (52). This has reduced instrument cost and increased availability (53). Operating on a similar measurement principle but on a very short range, trap-based systems have further reduced cost (54–57).

Current state of the art in lidar entomology

Entomological lidar systems have proven the capability to operate for multiple months and record long running dataset of insect activity at millisecond resolution (18). This in turns allows the study of seasonal trends and variation. Entomological lidar have also been used to study momentary changes such as the effect on insect activity from a solar eclipse (17). Such studies would be very resource intensive or even impossible to perform with conventional methods. However, they still have limited ability to provide species specific information (58). Multiple efforts to develop species specific classification systems have been pursued (37,38,59). For practical reasons, labelled data has been collected with close range laboratory equipment in controlled conditions, rather than long range lidars.

Some groups have used time signals as input to neural networks and applied sophisticated deep learning methods (60). This approach has the advantage of reducing the number of feature extraction algorithms needed but it can be difficult to transfer the methodology to other systems. Other groups have used numerically estimated physical features and compared more simple classification algorithms (38). An advantage of using physical, optical properties is that they can potentially be measured from tethered or even dead museum specimens and shared between different instruments (51,61–63).

A common limitation of previous lidar work has been the low number of species included in the studies and the limited ability to transfer classification models to the field. While it is impressive to distinguish a handful of species in controlled conditions, Insecta is one of the most species rich orders in the animal kingdom and there is a high risk of false positives.

While it remains challenging to prove that entomological lidar and similar systems can provide species specific measurements in the field, clustering efforts have been made. Recent studies have demonstrated the ability to cluster insects from two physical features into 4 distinct groups (64). While these groups were manually identified from features in the data, they showed different *diel* and seasonal activity distributions. Other groups used hierarchical clustering to identify the top 20 clusters in entomological lidar data collected over one week of measurements (17). Using the same clustering algorithm, another study used the elbow method to quantify the number of clusters found in a 500 m transect (65).

Light-matter interactions

Entomological lidar and similar optical sensors depend on the physical interaction between light and a target insect. Photons are emitted from an instrument, interact with an insect, and are captured by a detector. As these interactions are well described elsewhere in the literature (53,66,67), only a brief cover of the main light-matter interactions relevant for lidar entomology is provided below.

Scattering

Scattering causes photons to change propagation direction. This happen when light encounters inhomogeneity/interfaces of refractive indices in the tissue (according to Fressnell's equations). Insects are largely transparent in the near infrared and in general, forward scattering is the dominating interaction (63,68). However, for practical reasons, entomological lidar has the emitter and detector placed in close proximity and are therefore limited to detecting back-scattered light (53). Back-scattered light can be either due to single scatter interactions, where the photons bounce on the surface of the insects, or multiple scattering interactions where the photons bounce multiple times inside the insect.

Polarization

Light waves are composed of waves in both the electric and magnetic fields. The direction of the electric field component gives the polarization and light waves can contain multiple polarizations at once. For linearly polarized light, the electric field component of the wave oscillates in a single direction, whereas its direction is random for unpolarized light.

As light interacts with matter, there is a chance that it changes polarization. As the number of interactions increases, the probability of losing the original polarization increases. The rate of de-polarization can be described by a depolarization probability of the target tissue (69). We can, for example, assume that a shiny beetle reflects less de-polarized light than a furry bumblebee. I have not used polarization sensitive instrumentation in this work, but the polarization ratio of insects is commonly used by other groups as a feature for species classification (51,68,70). It has also been shown that the degree of linear polarization can be used to distinguish gravid, and non-gravid mosquitoes (71).

Absorption

In addition to scattering, light interacting with a medium can be absorbed and converted to thermal energy. The absorption rate varies with the wavelength and molecular composition of the insects but of primary interest for lidar entomologists is the melanin absorption spectra (72,73). By using multiple wavelengths with different absorption rates, the melanisation of the target insects can be estimated (39,74).

Interference

When multiple light waves with different phases interact, they produce a new wave which is the superposition of the original waves. If light travels through a thin film, such as an insect wing, it interferes with itself as it bounces back and forth inside the element. Depending on the wavelength of the light, and the thickness of the element, the interference can be either constructive or destructive. By comparing the reflections of multiple wavelengths, the thickness of the element can be estimated. This has recently allowed precise measurements of insect wing thickness (61–63,74–76).

Machine learning and ecology

Automated insect monitoring systems depend on automated analysis of the captured data. As the availability of machine learning models has increased in recent years, it has also become a more common tool in ecology and entomology (15,77,78). In short, machine learning models can be sorted into two categories, supervised and unsupervised models as illustrated in Figure 1.

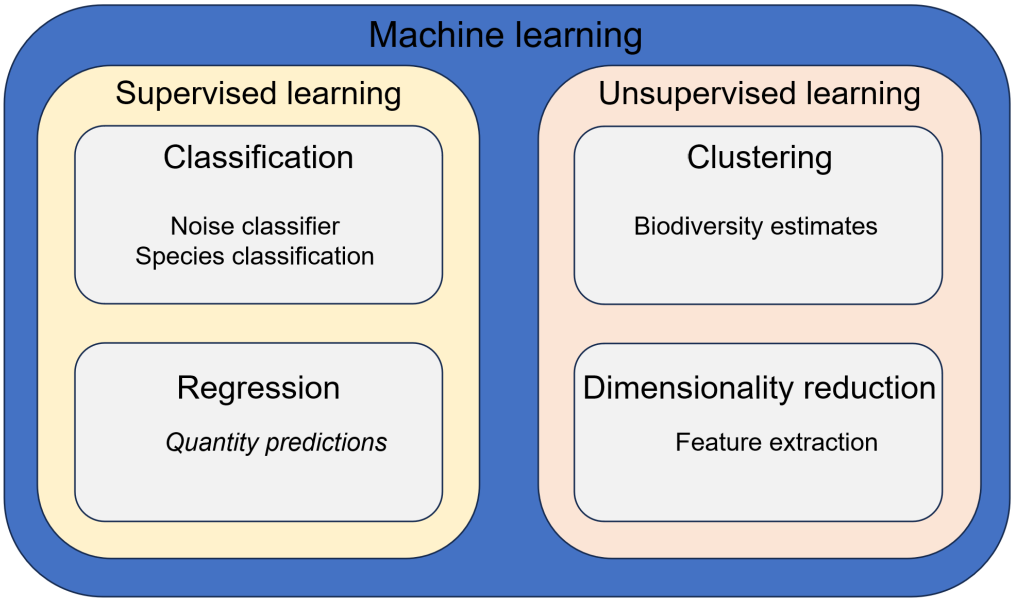


Figure 1: A schematic overview of machine learning algorithms and their applications in this thesis. Regression has not been used extensively in this work.

Supervised methods primarily include classification, where a model classifies a piece of data into two or more separate classes, and regression, where a model aims to predict a quantity, such as the weight or flight direction from the data. Classification models include segmentation models where an object such as an insect is found within, and isolated from, a larger image. Common for all supervised methods is that they require a labelled reference dataset, where each data point has a "true" reference value. The model is then trained to make predictions on the training data and its performance is judged on the agreement between predictions and the "true" reference values.

There are multiple metrics available for quantifying the performance of classification and regression models. A common method for classifiers is the sum of the number of correctly predicted targets (True positives) and others (True negatives) divided by the total size of the dataset. However, by investigating the distribution of wrongly classified targets (False negatives) and others (False positives), further insight can be gained. Supervised models can become very

powerful but collecting the labelled dataset is in general a time-consuming effort that requires human subject matter experts.

In ecology and particular entomology, machine learning classifiers are used for taxonomic identification, both on collected microscopy images from collected samples and in camera traps (20,77,79). Classifiers are able to combine multiple images of the same insect, recorded from different angles or with a different focus to improve the accuracy (79,80). Classifiers can also be used to identify the collected pollen from bees to identify food networks, recognize birds from audio recordings etc. (77). They are available to the public in smartphone apps, capable of identifying animals or plants from photos (81). In general, supervised models are limited by the availability of high quality labelled data. The amount of data needed to train a model increases with number of model parameters, i.e. its complexity.

Unsupervised methods are also affected by the amount of training data, but it does not require manual labelling. Both supervised models and clustering methods require pre-processing to unify the data before their application. A common pre-processing step is the dimensionality reduction, where the amount of data for each data point is unified and reduced to simplify the model. The variational auto encoder described in paper III is a typical example of an unsupervised, dimensionality reducing model. The clustering methods described in paper IV and V are also unsupervised methods.

Unsupervised methods are frequently used with DNA based methods, especially in fields such as mycology and soil science, where a large part of the genome is unmapped. By clustering the data, researchers can find common traits between clusters and find new patterns (16,82). While unsupervised methods do not require labelled training data, they in general have tuneable parameters that affects the number of clusters generated, or the similarity metric used compare the datapoints.

Understanding entomological lidar data

In my thesis, I have mainly worked with FaunaPhotonics´ "Volito" sensor. It is a close range, dual band entomological lidar system described in detail in paper II. An example of a insect recording from Volito sensor is shown in figure 2.

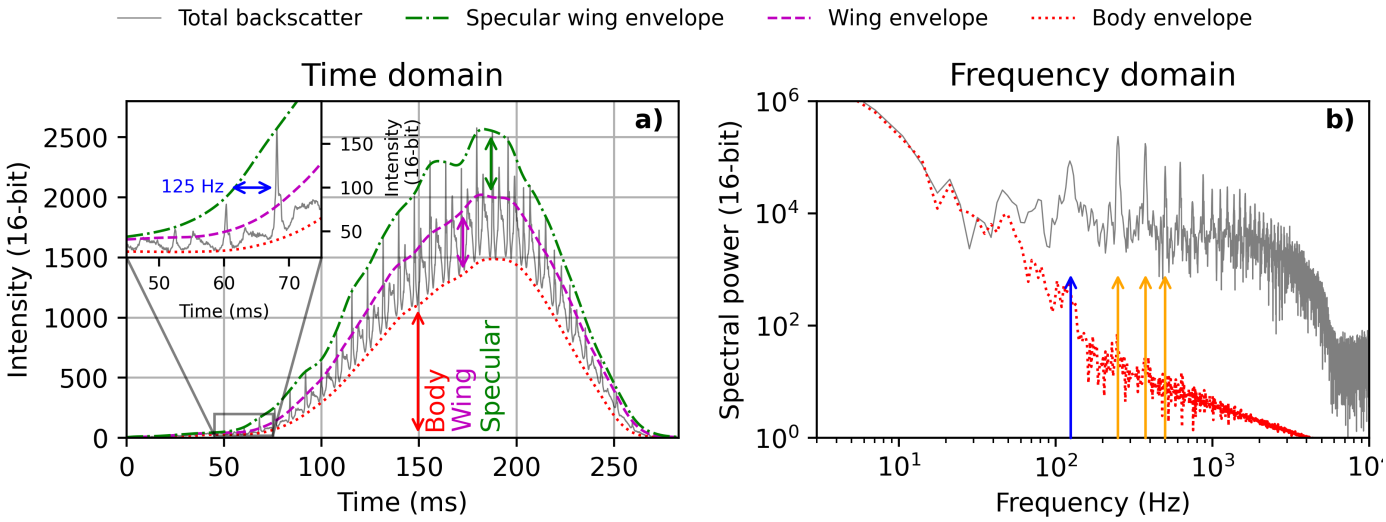


Figure 2: Example of an insect recording from the Volito sensor, adapted from paper V. a) Signal is recorded in time and the body, wing and specular wing magnitudes are estimated. b) Fourier transformation of the total backscatter and body envelope. The fundamental frequency is indicated in blue. The first three overtones are indicated in orange. The Volito has a bandwidth of 0-5 kHz. In this case the second harmonic is stronger than the fundamental. The reason for the limited magnitudes of the harmonics is that each one is convoluted by the body envelope (66,67).

In general, we assume that an insect recording has three frequency components in each wavelength band:

- A body signal
- A diffuse wing signal
- A specular wing signal

The contribution from the insect body is the result of single and multiple scattering events in the body of the insect. As this signal does not have any oscillatory properties, it can be estimated as the lower envelope of the signal. The diffuse wing signal is the result of scattering from scaled wings or veins in clear winged insects. As the projection, and thus the optical cross section, of the wings varies with their movement, the diffuse wing signal is seen as a periodic contribution on top of the body signal.

The specular wing signal, visible as rapid flashes in intensity with the same periodicity as the diffuse wing signal but with much shorter duration. This signal is caused by specular reflections when clear wings surface normal intercepts the midpoint between light source and detector. A mathematical formulation of these signals are provided as an appendix to paper V.

From the body, diffuse and specular wing signal components, we can estimate number of physical features. Of special relevance is the wingbeat frequency since it is closely linked to the anatomy of each species. While temperature dependent, it is relatively consistent within species and comparatively well described in literature (37,83–87). There is, however, a large overlap between species as most insects have wingbeat frequencies between 25 and 1000 Hz. The average spread within the 42 species used in paper V was 20 Hz ±10 % of the mean.

Classifying and clustering entomological lidar data

Extracted features in lidar data are caused by morphological and behavioural properties of insects and are in general consistent within species. Therefore, we can use such features for species classification and biodiversity estimates. However, as any object that traverses the beam will generate a signal, including rain, leaves, dust, etc. insect observations need to be isolated from recordings generated by other objects. The first step in this noise filtering is often to remove all observations without identifiable wingbeat patterns. In my work, I have also used a neural network, trained on manually labelled observations, to further remove non insect observations.

There are many different approaches to develop a classifier to identify species from others and a fundamental task for all of them is to draw boundaries in the feature space, minimizing the overlap between target and negative classes. In contrast, the task of estimating biodiversity is less well defined, not only for entomological lidar but also in conventional ecology.

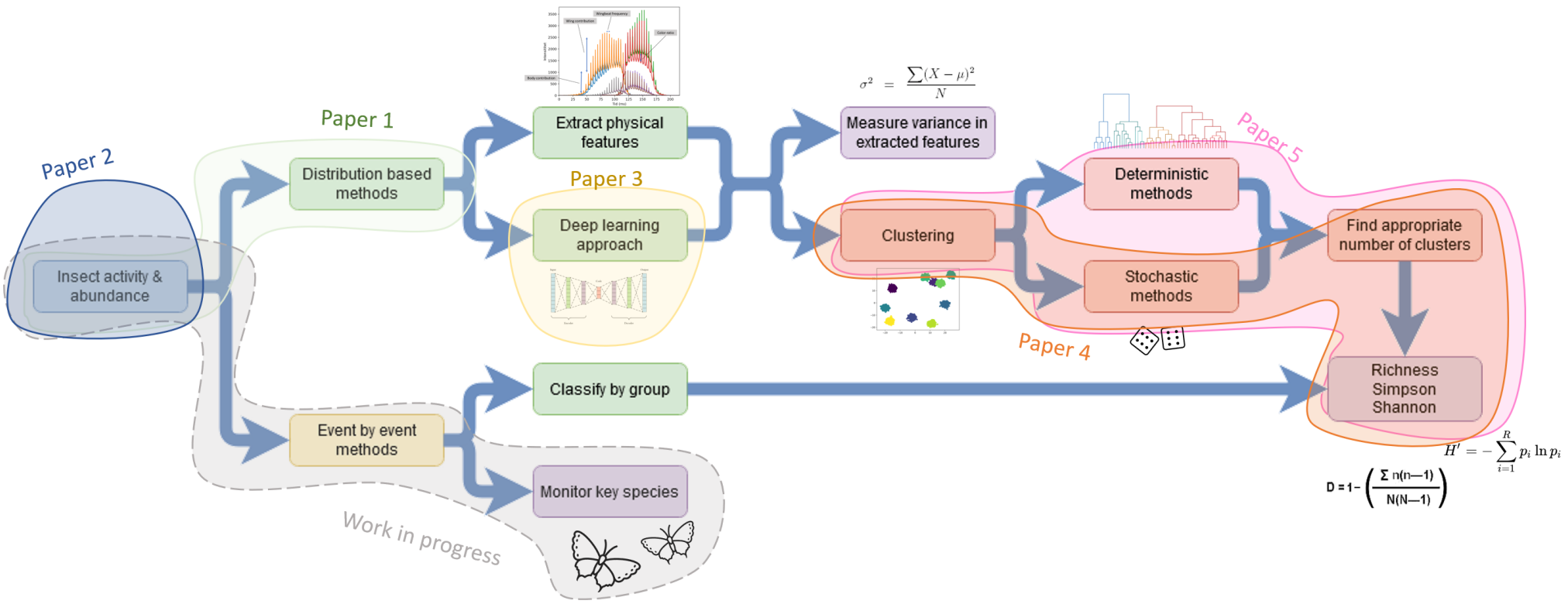


Figure 3: A schematic overview of possible paths to measure biodiversity with photonic sensors and similar methods. The paths explored in this thesis are indicated. The work in progress refers to laboratory and field experiments with honeybees and bumblebees not ready to be included in the thesis.

One approach is to mimic the conventional monitoring method of identifying all observations to species level. This would allow for any conventional metric to be applied to optically recorded data. However, the task of collecting data and training classifiers for every species in the world seems unrealistic in the near and even in the far future. As a starting point, the taxonomic level could be to identify broad groups, such as family, or even order. However, even this remains challenging.

Another approach is to focus on indicator species. Indicator species are commonly used in conventional monitoring methods (88). However, it is an approach that is poorly suited for entomological lidar where any rare species is outweighed by a large factor by common species such as flies and there is a high risk that false positives will influence the results.

Alternatively, one can use biodiversity estimates such as species richness, or the Shannon and Simpson indices. These estimates are based on the number of unique species or registrations and their relative proportions but do not include the genetic relationship between the species. It is therefore not necessary to identify which species are present, just how many of each kind. This is the main approach I have pursued in this work, since it allows a flexible statistical approach to clustering, while it is conceptually similar to methods used with conventional monitoring methods.

A final approach, not explored in this thesis, is to abstain from direct translations of methods used with conventional monitoring methods. Rather than classifying or clustering individual observations, the general distribution of features can be quantified. For example, the variance of recorded wingbeat frequencies could be used as a metric. Such a metric would not have any obvious equivalents in conventional monitoring. More complex metrics such as entropy measurements across a multi-dimensional feature space is also likely to perform well. However, I have avoided these approaches due to time limitations. Additionally, their lack of analogies to conventional monitoring might render them more difficult to get adopted by the scientific and general community.

Research objectives

The goal of this project was to develop and evaluate the use of automated optical sensors in insect monitoring. While entomological lidars have been reported in multiple countries, there has been few comparative studies, relating their results to conventional ground truthed data. Most of the previous work have been focused on shorter measurement campaigns rather than extended, seasonal, monitoring schemes.

The first task focused on the possibility to use distributed sensors to quantify insect biodiversity. We aimed to focus on a simple metric such as richness, and the related Shannon and Simpson's biodiversity indices. The goal was to develop a tool for researchers and ecologists, enabling them to answer relatively simple questions such as "does a flower strip increase the biodiversity in a field or not?", "at what time of the year is the biodiversity at its peak?", "does the application of insecticide reduce the biodiversity within a field?" and in the future, more complex questions such as "is this action an effective way to increase biodiversity on a landscape level"? These goals have been pursued in paper III-V.

The second task focused more specifically on bee monitoring. Managed bees are important actors in the agricultural landscape but there is concern that competition between managed honeybees and wild bees is partly to blame for the decline of wild bee communities. The decline of native bee communities has received a lot of attention from the public, which has engaged with bee hotels and reserved meadow areas in private and public gardens. As honeybees and bumblebees are flight active, relatively large insects that are commercially available and easy to manage, they were deemed good targets for an initial, proof of concept classification algorithm. The goal was to develop a honeybee counter which would be a useful tool for studying competition effect and investigate foraging range and pollination strategies.´

This work has been pursued in paper I and in an additional experiment. In the latter, a honeybee and bumblebee hive were separately installed in a large plant tunnel. A food source was placed in the far end of the tunnel along with three Volito sensors, which recorded the bees as they flew back and forth between the hive and the food source. To validate a future classifier, a field experiment was conducted.

Three different phacelia flower strips were selected in areas without any registred honeybee hives. At each site, three beehives were installed in cages, allowing us to modify the honeybee access to the flower strip. At 50, 100 and 150 m distance from the hives, along the flower strips, Volito sensors were installed. The experiment ran over two weeks and the honeybee cages were opened and closed every two days. The experiment was validated by manual observations, were all honey- and bumblebees entering a 1 m^2 square in front of the sensor in a 10 min period was manually counted by visual inspection. This data is still being treated.

Paper abstracts

Paper I: Scheimpflug lidar range profiling of bee activity patterns and spatial distribution

In this work, we used a long range entomological lidar to study the foraging behaviour of managed honeybees in a white clover field. Commercial honeybees are frequently used for pollination of fruit trees and vegetables as well as arable crops and while wild bees are less important in such settings, they are essential for the pollination of wild plants. However, honeybees may also contribute to declines of wild bee communities through competition and spread of diseases. To reduce the impact of commercially deployed honeybees on the wild insect communities, we need to study the temporal and spatial behaviour of both managed and wild bees. As such studies are complicated to conduct with conventional monitoring methods, we evaluated the performance of an entomological long range lidar.

The lidar monitored a 1 km long transect and passed close to a cluster of beehives at ca 180 m distance. The insect activity was recorded over three days and the distribution of insects in the field was mapped. In total 566 609 insect observations were recorded, and the spatial distribution was separated into three groups, with two centred on the beehives. Using these distributions, the temporal activity and foraging range was estimated. In addition to the lidar measurements, we conducted ground truthing by performing transect walks. We also used hive scales to monitor the weight change of the hives, from which the flight activity could be estimated. The measured honeybee activity in the lidar was well correlated with both the transect walks and hive scales.

The study showed that by designing the measurements around a point source, or in the future, an attractor such as a food source, behavioural studies can be conducted using long range entomological lidar without the classification of individual insect observations. Additionally, some insight into the three-dimensional distribution of honeybees was gained by altering the beam height between a lower and a higher transect, which revealed a funnel like distribution around the hives.

Relationship to state of the art

While the very first application of entomological lidar also focused on honeybee monitoring (51), it has not been repeated since. The publication included the first attempt to use lidar to estimate foraging ranges. Additionally, it is, to my knowledge the first-time insect abundance measures from lidar were correlated with manual observations of honeybees.

Paper II: Automating insect monitoring using unsupervised near-infrared sensors

This paper describes the development and deployment of FaunaPhotonics novel short range sensor, later named “Volito” which was used in the following publications. While the long range entomological lidars can provide a large number of observations, the field installation is cumbersome and requires highly trained operators. Additionally, the instrumentation requires close supervision due to its sensitive alignment and eye-safety risks.

To allow for continuous monitoring in remote areas, and repeated landscape studies, a smaller and more robust sensor was needed. The Volito is a close range field monitoring device powered by a solar panel and equipped with a cellular connection to allow remote monitoring and real time data acquisition. The sensor automatically extracts insect observations from the raw data stream and transmits these to a cloud database for further processing.

The Volito uses a dual band infrared LED array and the light is emitted in a wide angle to improve eye safety and maximize the near field measurement volume. The effective measurement volume varies with the size of the target but using a custom-built mapping robot, it was measured to between 5 and 100 liters depending on the configuration of the sensor and the size of the target insect.

Six sensors were deployed in an oilseed rape field over four weeks along with water traps to evaluate the performance against conventional monitoring methods in commercial crops. The water traps were emptied daily and the aggregated insect counts in sensors and traps were comparable (Spearman rank correlation of 0.6). The main purpose of the paper was to describe the sensor in detail, to have a common reference and allow further publications to refer to technical details to this paper.

Relationship to state of the art

While entomological lidar has become increasingly common in recent years (65,89–93), they have been cumbersome to deploy. This paper describes the first photonic sensor, measuring an open-air volume, capable of autonomous long-term monitoring in the field. It includes the first characterizing of the probe volume using a sphere dropping robot. It also covers the first long term deployment and comparison between photonic sensors and yellow water traps in an agricultural setting.

Paper III: Dynamic β-VAE for quantifying biodiversity by clustering optically recorded insect signals

In this paper, we evaluated a deep learning approach to extract a low dimensional feature representation of the insect observations to allow for later biodiversity clustering. A parameter space becomes increasingly empty with increasing dimensions, commonly known as the "curse of dimensionality". Regardless of clustering method used, it is thus advantageous to represent your data with as few features as possible, while still capturing sufficient information to allow for the downstream tasks.

In entomological lidar, a low dimensional representation is typically constituted by estimated physical features such as a wingbeat frequency, melanization, body wing ratio etc. Estimating these features is a difficult and potentially error prone task where the algorithms are sensitive to outliers. An alternative is to use the wingbeat frequency power spectra but in order to achieve a sufficient spectral resolution, a comparatively large number bins are needed.

In this work, we developed a dynamic variational auto encoder (VAE) to automatically transform a high dimensional frequency representation of the data into a two-dimensional and well regularized feature space. As phylogenetic insect groups are clustered together, the approach was deemed suitable for later richness estimations. β-VAE's introduces a scaling term, β, to balance the regularization and reconstruction losses. In this work, we introduced a dynamic self-adjustment of β which greatly improved model stability and results. This approach made it possible to achieve well regularized latent representation while also retaining high quality reconstructions. Unlabeled data recorded in the field was used to train the model.

The performance of the model was based on the ability to cluster similar species together, using the ARI and AMI score. This required labelled data and we used insect observations recorded one species at the time in flight cages. The β-VAE was compared to alternative more simple methods, including PCA and hierarchical clustering. The β-VAE achieved two to three times better results than the conventional methods. To further improve the ability to cluster single species together, we tested to include a small subset of labelled data in the training. In this experiment, an additional loss term based on the intra and inter cluster distances between the labelled data points were added. This yielded an improvement of ca 10% as compared to the fully unsupervised β-VAE.

While exploratory, this work showed that phylogenetically similar species in general share similar properties when recorded by entomological lidar. This verified fundamental assumption for future richness estimations, regardless of method used for feature extraction, or richness estimation.

Relationship to state of the art

This publication describes the first application of a variational auto encoder on photonic insect data. While regular auto encoders have previously been used for feature extraction before classification. (94), they are unsuitable for unsupervised clustering due to their highly irregular latent spaces. The paper show that using the proposed models, the results improve upon the methods used in previous state of the art (65) by ca. 100 – 300%. Additionally, it introduces dynamic scaling of the β-term which improved the results ca 80% over the regular VAE.

Paper IV: Automating an insect biodiversity metric using distributed optical sensors: an evaluation across Kansas, USA cropping systems

This manuscript describes the first experiment conducted to compare biodiversity estimates between sensors and conventional methods. Financed and partly planned by General Mills, 20 fields covering 6 different crops and various regenerative farming practices were investigated. The fields were sampled using sweep netting, Malaise traps and Volito sensors twice during the growth season. Malaise traps were open for 24 hours and the sensors were active for three days each sample period. Collected specimen were in general identified to species level. Additionally, ecosystem indicators such as the predation rates were measured once per field.

The experiment suffered from some limitations. Firstly, sensors were installed close to the growing crop, which generated large amounts of noise in the data. Secondly, the sampling was not done simultaneously for all methods. It is therefore likely that weather effects influenced the data and the measured insect abundance was uncorrelated across all methods. We used one third of the data to optimize the hyperparameters of the DBSCAN clustering algorithm. The number of clusters, and their relative distribution, generated by the algorithm was compared to the results from the conventional measurements of species richness Shannon and Simpsons biodiversity indices.

Simpson and Shannons biodiversity indices were largely uncorrelated across all methods. Species richness was weakly correlated between the two conventional methods (Spearman R: 0.36, $p<0.05$, N=40). The correlations between the number of clusters generated by the algorithm and both individual sampling methods were high (Spearman R: 0.48 and 0.52, $p<0.05$, N=40). The correlation between the number of clusters generated by the algorithm and the combined richness of the individual sampling methods was also high (Spearman R: 0.55, $p<0.05$, N=40) In this project, I got involved after the experiment was completed. My task was to take over the data analysis and write a first, technical draft of the manuscript.

Relationship to state of the art

This manuscript describes the first attempt to use photonic sensor for insect biodiversity measurements by comparing the results to conventional monitoring methods. Previous work has aimed to quantify richness in entomological lidar data but haven't been ground truthed or compared with conventional monitoring methods (65). This manuscript contains the first proof of concept that photonic sensors can quantify insect diversity.

Paper V: Photonic sensors for comparative insect abundance and diversity in distinct habitats

The manuscript describes the development and testing of two different approaches to quantifying species richness with photonic sensors. The models were developed and optimized on labelled data, collected in flight cages during controlled conditions, and evaluated in the field. The field tests included 5 different locations in southern Sweden and Denmark where sensors and Malaise traps were deployed simultaneously from March to November. Traps were emptied weekly and caught specimen were identified to family level. In total, 9503 insects were identified.

The labelled dataset used for development was composed of 42 species, recorded one species at the time in flight cages. Using this library, we assembled arbitrary compositions of data with known diversity. We developed two tests to evaluate the models. The first test quantified the correlation between the number of clusters generated by the algorithms against the number of species present in the data. The second tests investigated the number of clusters generated from increasing number of observations from a single species.

We investigated the DBSCAN algorithm, also used in paper IV, as well as a hierarchal clustering algorithm (HCA). The DBSCAN algorithm operated on a low dimensional feature representation of the data while the HCA operated on the high dimensional wingbeat frequency power spectra.

A parameter sweep was done to optimize hyperparameters on the labelled data. The best performing configurations were then tested on the data collected in the field. Both methods showed good correlations with the Malaise trap data (Pearson R: 0.54, 0.67, $p<0.05$, N=40, for DBSCAN and HCA respectively) but the relative sizes of the clusters were not well correlated with the distribution of families in the traps.

Relationship to state of the art

This manuscript describes the first efforts to validate the biodiversity metrics achieved from photonic sensors on a fully independent dataset. It uses a unique collection of labelled insect data, spanning 42 species to develop the methodology. This is a 400% increase over the largest previously reported database of photonic insect signals (59,95). Additionally, it covers the first long term comparison between photonic sensors and Malaise traps for both abundance and diversity.

Conclusions and perspectives for further research

This project has aimed to develop a proof of concept for the use of entomological lidar in entomological and ecological studies. To tackle the decline of insect diversity and abundance, novel monitoring methods are needed. While the insight gained by regular field visits by skilled experts cannot be replaced by machines, the automated methods can complement the conventional monitoring methods and provide data in real time as shown in paper II, IV and V.

At the start of the project, there was a handful of entomological lidars available globally and they in general required technical experts to set up and utilize. This project included the very first installation of FaunaPhotonics "Volito" sensor. At the end of this PhD project, FaunaPhotonics has deployed hundreds of sensors across in Europe and America. The cost is still high compared to conventional equipment and access for researchers to the technology is still limited but the methodology is gaining traction.

The ability to accurately identify single observations to species level is still limited and difficult to validate in the field. The experiment with honeybees in flower strips aimed to collect a ground truthed-dataset for classifier development. Four persons working in the field for two weeks yielded ca 400 manually observed ten-minute time periods. This work is not ready for inclusion in this thesis, but it illustrates the work necessary to validate the performance of a single classifier. While scaling to more species will be a challenge, I believe "just" an automated bee classifier would be a great asset for ecologists worldwide.

This PhD project yielded the first validated proof-of-concept of biodiversity monitoring using photonic sensors, but a lot of work remains. I have investigated two clustering algorithms but there are many approaches available. Future work could combine the deep learning feature extraction from paper III with the validation approach developed in paper V. Additionally, more sites with where the species richness and abundance are more uncorrelated should be investigated.

Although we have not developed a new golden standard for biodiversity monitoring with automated sensors yet, or a high performing bee classifier, I do believe this project collected the data necessary for this future work. The tools and concepts explored in this work are transferrable to both the acoustic and radar domains. A big advantage of the automated methods is the ability to scale, and I hope we will see large regional networks of live biodiversity sensors in the future.

References

1. Weisser WW, Siemann E. Insects and ecosystem function. Vol. 173. Springer Science & Business Media; 2013.
2. LOSEY JE, VAUGHAN M. The Economic Value of Ecological Services Provided by Insects. Bioscience. 2006;56(4):311.
3. Loreau M, Hector A, Isbell F. The ecological and societal consequences of biodiversity loss. John Wiley & Sons; 2022.
4. Organization WH. A global brief on vector-borne diseases [Internet]. World Health Organization; 2014 [cited 2023 Apr 11]. Available from: www.who.int
5. Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, et al. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS One. 2017;12(10).
6. Wagner DL, Grames EM, Forister ML, Berenbaum MR, Stopak D. Insect decline in the Anthropocene: Death by a thousand cuts. Proc Natl Acad Sci U S A [Internet]. 2021 Jan 11 [cited 2023 Apr 11];118(2). Available from: https://doi.org/10.1073/pnas.2023989118
7. Neff F, Korner-Nievergelt F, Rey E, Albrecht M, Bollmann K, Cahenzli F, et al. Different roles of concurring climate and regional land-use changes in past 40 years' insect trends. Nat Commun [Internet]. 2022;13(1):7611. Available from: https://doi.org/10.1038/s41467-022-35223-3
8. Seibold S, Gossner MM, Simons NK, Blüthgen N, Müller J, Ambarlı D, et al. Arthropod decline in grasslands and forests is associated with landscape-level drivers. Vol. 574, Nature. 2019. 671–674 p.
9. Karlsson D, Hartop E, Forshage M, Jaschhof M, Ronquist F. The Swedish Malaise Trap Project: A 15 Year Retrospective on a Countrywide Insect Inventory. Biodivers Data J [Internet]. 21AD Jan 2020;8:e47255. Available from: https://doi.org/10.3897/BDJ.8.e47255
10. Montgomery GA, Belitz MW, Guralnick RP, Tingley MW. Standards and Best Practices for Monitoring and Benchmarking Insects. Vol. 8, Frontiers in Ecology and Evolution. Frontiers Media S.A.; 2021. p. 579193.
11. Morris RF. Sampling insect populations. Annu Rev Entomol. 1960;5(1):243–64.
12. Pollard E. A method for assessing changes in the abundance of butterflies. Biol Conserv. 1977;12(2):115–34.

13. Strickland AH. Sampling crop pests and their hosts. Annu Rev Entomol. 1961;6(1):201–20.
14. Karlsson D, Hartop E, Forshage M, Jaschhof M, Ronquist F. The Swedish Malaise Trap Project: A 15 Year Retrospective on a Countrywide Insect Inventory. Biodivers Data J [Internet]. 2020 [cited 2023 Feb 17];8:47255. Available from: /pmc/articles/PMC6987249/
15. van Klink R, August T, Bas Y, Bodesheim P, Bonn A, Fossøy F, et al. Emerging technologies revolutionise insect ecology and monitoring. Trends Ecol Evol. 2022;37(10):872–85.
16. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol [Internet]. 2012 Apr 1 [cited 2023 Feb 17];21(8):2045–50. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-294X.2012.05470.x
17. Brydegaard M, Jansson S, Malmqvist E, Mlacha Y, Gebru A, Okumu F, et al. Lidar reveals activity anomaly of malaria vectors during pan-African eclipse. Sci Adv. 2020 May 13;6.
18. Genoud AP, Williams GM, Thomas BP. Continuous monitoring of aerial density and circadian rhythms of flying insects in a semi-urban environment. PLoS One [Internet]. 2021;16(11 November):1–16. Available from: http://dx.doi.org/10.1371/journal.pone.0260167
19. Marić M, Orović I, Stanković S. Compressive Sensing based image processing in TrapView pest monitoring system. In: 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE; 2016. p. 508–12.
20. Preti M, Verheggen F, Angeli S. Insect pest monitoring with camera-equipped traps: strengths and limitations. J Pest Sci (2004) [Internet]. 2021 Mar 1 [cited 2023 Jun 9];94(2):203–17. Available from: https://link.springer.com/article/10.1007/s10340-020-01309-4
21. Potamitis I, Eliopoulos P, Rigakis I. Automated Remote Insect Surveillance at a Global Scale and the Internet of Things. Robotics. 2017;6(3).
22. Ascolese R, Gargiulo S, Pace R, Nappa P, Griffo R, Nugnes F, et al. E-traps: A valuable monitoring tool to be improved. EPPO Bulletin [Internet]. 2022 Apr 1 [cited 2023 Jun 8];52(1):175–84. Available from: https://onlinelibrary-wiley-com.ep.fjernadgang.kb.dk/doi/full/10.1111/epp.12838

23. Butail S, Manoukis N, Diallo M, Ribeiro JM, Lehmann T, Paley DA. Reconstructing the flight kinematics of swarming and mating in wild mosquitoes. J R Soc Interface. 2012;9(75):2624–38.
24. Hall ML, Gleave K, Hughes A, McCall PJ, Towers CE, Towers DP. The application of digital holography for accurate three-dimensional localisation of mosquito-bednet interaction. Light: Advanced Manufacturing. 2022;3(3):1.
25. Saleh BEA, Teich MC. Fundamentals of Photonics. 1991 Aug 14 [cited 2023 Jun 8]; Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/0471213748
26. Bjerge K, Mann HMR, Høye TT. Real-time insect tracking and monitoring with computer vision and deep learning. Remote Sens Ecol Conserv [Internet]. 2022 Jun 1 [cited 2023 Feb 20];8(3):315–27. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/rse2.245
27. diopsis.eu | An advanced, modern system for photographing, recognizing and monitoring insects in a fully automated way [Internet]. [cited 2023 Jun 8]. Available from: https://diopsis.eu/en/
28. Biodiversity sensor | Syngenta [Internet]. [cited 2023 Jun 8]. Available from: https://www.syngentagroup.com/en/sustainability/biodiversity-sensor
29. Mankin RW, Hagstrum DW, Smith MT, Roda AL, Kairo MTK. Perspective and promise: A century of insect acoustic detection and monitoring. American Entomologist. 2011;57(1):30–44.
30. Sinka ME, Zilli D, Li Y, Kiskin I, Msaky D, Kihonda J, et al. HumBug – An Acoustic Mosquito Monitoring Tool for use on budget smartphones. Methods Ecol Evol [Internet]. 2021 Oct 1 [cited 2023 Feb 20];12(10):1848–59. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13663
31. Jeliazkov A, Bas Y, Kerbiriou C, Julien JF, Penone C, Le Viol I. Large-scale semi-automated acoustic monitoring allows to detect temporal decline of bush-crickets. Glob Ecol Conserv. 2016 Apr 1;6:208–18.
32. Peer-reviewed NOT. TreeVibes: Modern tools for global monitoring of trees against borers Iraklis Rigakis, 1 Ilyas Potamitis, 2 Nicolaos-Alexandros Tatlas,. 2020;(November):1–16.
33. Hagstrum DW, Flinn PW, Shuman D. Automated Monitoring Using Acoustical Sensors for Insects in Farm-Stored Wheat. J Econ Entomol [Internet]. 1996 Feb 1 [cited 2023 Jun 8];89(1):211–7. Available from: https://dx.doi.org/10.1093/jee/89.1.211

34. Eliopoulos PA, Potamitis I, Kontodimas DC, Givropoulou EG. Detection of Adult Beetles Inside the Stored Wheat Mass Based on Their Acoustic Emissions. J Econ Entomol. 2015 Dec 1;108(6):2808–14.
35. Hu C, Kong S, Wang R, Long T, Fu X. Identification of Migratory Insects from their Physical Features using a Decision-Tree Support Vector Machine and its Application to Radar Entomology. Scientific Reports 2018 8:1 [Internet]. 2018 Apr 3 [cited 2023 Jun 8];8(1):1–11. Available from: https://www.nature.com/articles/s41598-018-23825-1
36. Gebru AK, Brydegaard M, Rohwer E, Neethling P. Probing insect backscatter cross-section and melanization using kHz optical remote detection system. 2016;11(1):997504. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2236010
37. Kirkeby C, Rydhmer K, Cook SM, Strand A, Torrance MT, Swain JL, et al. Advances in automatic identification of flying insects using optical sensors and machine learning. Sci Rep. 2021;11(1):1555.
38. Genoud AP, Gao Y, Williams GM, Thomas BP. A comparison of supervised machine learning algorithms for mosquito identification from backscattered optical signals. Ecol Inform. 2020 Jul 1;58:101090.
39. Gebru A, Brydegaard M, Rohwer E, Neethling P. Probing insect backscatter cross section and melanization using kHz optical remote detection system. J Appl Remote Sens. 2017;11(1):16015.
40. Friis HT. Radar reflections from the lower atmosphere. Proceedings of the Institute of Radio Engineers. 1947;35(5):494–5.
41. Crawford AB. Radar reflections in the lower atmosphere. Proceedings of the Institute of Radio Engineers. 1949;37(4):404–5.
42. Schaefer GW. Radar studies of locust, moth and butterfly migration in the Sahara. Proc Royal Entomol Soc Lond C. 1969;34(33):39–40.
43. Drake VA, Reynolds DR. Radar entomology: observing insect flight and migration. Cabi; 2012.
44. Hao Z, Drake VA, Taylor JR, Warrant E. Insect target classes discerned from entomological radar data. Remote Sens (Basel). 2020;12(4):1–18.
45. Lin TY, Winner K, Bernstein G, Mittal A, Dokter AM, Horton KG, et al. MistNet: Measuring historical bird migration in the US using archived weather radar data and convolutional neural networks. Methods Ecol Evol [Internet]. 2019 Nov 1 [cited 2023 Jun

29];10(11):1908–22. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13280

46. Shamoun-Baranes J, Bauer S, Chapman JW, Desmet P, Dokter AM, Farnsworth A, et al. Weather radars' role in biodiversity monitoring. Science (1979) [Internet]. 2021 Apr 16 [cited 2023 Jun 29];372(6539):248. Available from: https://www.science.org/doi/10.1126/science.abi4680
47. Riley JR, Smith AD, Reynolds DR, Edwards AS, Osborne JL, Williams IH, et al. Tracking bees with harmonic radar. Nature [Internet]. 1996;379(6560):29–30. Available from: https://doi.org/10.1038/379029b0
48. Osborne JL, Clark SJ, Morris RJ, Williams IH, Riley JR, Smith AD, et al. A landscape-scale study of bumble bee foraging range and constancy, using harmonic radar. Journal of Applied Ecology. 1999;36(4):519–33.
49. Toyoshima M. Recent Trends in Space Laser Communications for Small Satellites and Constellations. Journal of Lightwave Technology. 2021 Feb 1;39(3):693–9.
50. Larsen EB. Studies on the activity of insects. III. Activity and migration of Plusia gamma. Bio-logiske Meddelelser. 1949;21:1–32.
51. Shaw JA, Seldomridge NL, Dunkle DL, Nugent PW, Spangler LH, Bromenshenk JJ, et al. Polarization lidar measurements of honey bees in flight for locating land mines. Opt Express. 2005;13(15):5853–63.
52. Brydegaard M, Malmqvist E, Jansson S, Larsson J, Török S, Zhao G. The Scheimpflug lidar method. https://doi.org/101117/122272939 [Internet]. 2017 Aug 30 [cited 2023 Jun 29];10406:104–20. Available from: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10406/104060I/The-Scheimpflug-lidar-method/10.1117/12.2272939.full
53. Brydegaard M, Svanberg S. Photonic Monitoring of Atmospheric and Aquatic Fauna. Laser Photon Rev. 2018 Dec 1;12(12).
54. Chen Y, Why A, Batista G, Mafra-Neto A, Keogh E. Flying insect classification with inexpensive sensors. J Insect Behav. 2014;27(5):657–77.
55. Potamitis I, Eliopoulos P, Rigakis I. Automated remote insect surveillance at a global scale and the internet of things. Robotics. 2017;6(3):1–14.
56. Potamitis I, Rigakis I, Tatlas NA. Automated surveillance of fruit flies. Sensors (Switzerland). 2017;17(1).
57. Johnson BJ, Weber M, Mohammad Al-Amin H, Geier M, Devine GJ. Automated differentiation of mixed populations of free-flying mosquitoes under semi-field

conditions. 2023 Jun 16 [cited 2023 Jun 29]; Available from: https://www.researchsquare.com

58. Bick E, Sigsgaard L, Torrance MT, Helmreich S, Still L, Beck B, et al. Dynamics of pollen beetle (Brassicogethes aeneus) immigration and colonisation of oilseed rape (Brassica napus) in Europe. Pest Manag Sci [Internet]. 2023 May 14 [cited 2023 Jun 16]; Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ps.7538
59. Silva DF, Souza VMA, Ellis DPW, Keogh EJ, Batista GE. Exploring low cost laser sensors to identify flying insect species. J Intell Robot Syst. 2015;80(1):313–30.
60. Qi Y, Cinar GT, Souza VMA, Batista GEAPA, Wang Y, Principe JC. Effective insect recognition using a stacked autoencoder with maximum correntropy criterion. Proceedings of the International Joint Conference on Neural Networks. 2015;2015-Septe.
61. Li M, Runemark A, Guilcher N, Hernandez J, Rota J, Brydegaard M. Feasibility of Insect Identification Based on Spectral Fringes Produced by Clear Wings. IEEE Journal of Selected Topics in Quantum Electronics. 2023;29(4).
62. Li M, Seinsche C, Jansson S, Hernandez J, Rota J, Warrant E, et al. Potential for identification of wild night-flying moths by remote infrared microscopy. J R Soc Interface [Internet]. 2022 [cited 2023 Jun 13];19(191). Available from: https://royalsocietypublishing.org/doi/10.1098/rsif.2022.0256
63. Li M, Jansson S, Runemark A, Peterson J, Kirkeby CT, Jönsson AM, et al. Bark beetles as lidar targets and prospects of photonic surveillance. J Biophotonics. 2020;(January 2021).
64. Genoud AP, Williams GM, Thomas BP. Continuous monitoring of aerial density and circadian rhythms of flying insects in a semi-urban environment. PLoS One [Internet]. 2021 Nov 18;16(11):e0260167. Available from: https://doi.org/10.1371/journal.pone.0260167
65. Kouakou BK, Jansson S, Brydegaard M, Zoueu JT. Entomological Scheimpflug lidar for estimating unique insect classes in-situ field test from Ivory Coast. OSA Contin [Internet]. 2020 Sep;3(9):2362–71. Available from: http://www.osapublishing.org/osac/abstract.cfm?URI=osac-3-9-2362
66. Jansson S. Entomological lidar : target characterization and field applications. Lund: Division of Combustion Physics, Department of Physics, Lund University; 2020.
67. Malmqvist E. From Fauna to Flames : remote sensing with Scheimpflug-Lidar [PhD thesis]. [Lund]: Lund University; 2019.

68. Jansson S, Atkinson P, Ignell R, Brydegaard M. First polarimetric investigation of malaria mosquitoes as lidar targets. IEEE Journal of Selected Topics in Quantum Electronics. 2018;25(1):1–10.
69. Jacques SL. Optical properties of biological tissues: a review. Phys Med Biol. 2013;58(11):R37.
70. Zhu S, Malmqvist E, Li W, Jansson S, Li Y. Insect abundance over Chinese rice fields in relation to environmental parameters , studied with a polarization - sensitive CW near - IR lidar system. Applied Physics B. 2017;1–11.
71. Genoud AP, Gao Y, Williams GM, Thomas BP. Identification of gravid mosquitoes from changes in spectral and polarimetric backscatter cross sections. J Biophotonics. 2019;12(10):e201900123.
72. Wolbarsht ML, Walsh AW, George G. Melanin, a unique biological absorber. Appl Opt. 1981;20(13):2184–6.
73. Sardar DK, Mayo ML, Glickman RD. Optical characterization of melanin. J Biomed Opt. 2001;6(4):404–11.
74. Gebru A, Jansson S, Ignell R, Kirkeby C, Prangsma JC, Brydegaard M. Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight. J Biophotonics. 2018;11(8).
75. Müller L, Li M, Månefjord H, Salvador J, Reistad N, Hernandez J, et al. Remote Nanoscopy with Infrared Elastic Hyperspectral Lidar. Advanced Science [Internet]. 2023 May 1 [cited 2023 Jun 13];10(15):2207110. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/advs.202207110
76. Månefjord H, Li M, Brackmann C, Reistad N, Runemark A, Rota J, et al. A biophotonic platform for quantitative analysis in the spatial, spectral, polarimetric, and goniometric domains. Review of Scientific Instruments [Internet]. 2022 Nov 1 [cited 2023 Jun 13];93(11):113709. Available from: /aip/rsi/article/93/11/113709/2849424/A-biophotonic-platform-for-quantitative-analysis
77. Pichler M, Hartig F. Machine learning and deep learning—A review for ecologists. Methods Ecol Evol [Internet]. 2023 Apr 1 [cited 2023 Jun 9];14(4):994–1016. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.14061
78. Høye TT, Ärje J, Bjerge K, Hansen OLP, Iosifidis A, Leese F, et al. Deep learning and computer vision will transform entomology. Proc Natl Acad Sci U S A [Internet]. 2021 Jan 11 [cited 2023 Jun 9];118(2):e2002545117. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.2002545117

79. Wührl L, Pylatiuk C, Giersch M, Lapp F, von Rintelen T, Balke M, et al. DiversityScanner: Robotic handling of small invertebrates with machine learning methods. Mol Ecol Resour. 2022;22(4):1626–38.
80. Ärje J, Melvad C, Jeppesen MR, Madsen SA, Raitoharju J, Rasmussen MS, et al. Automatic image-based identification and biomass estimation of invertebrates. Methods Ecol Evol. 2020;11(8):922–31.
81. Horn G Van, Aodha O Mac, Song Y, Cui Y, Sun C, Shepard A, et al. The iNaturalist Species Classification and Detection Dataset. [cited 2023 Jun 9]; Available from: www.inaturalist.org
82. Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, et al. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 2017 8:1 [Internet]. 2017 Oct 30 [cited 2023 Jun 29];8(1):1–11. Available from: https://www.nature.com/articles/s41467-017-01312-x
83. Potamitis I, Rigakis I. Large aperture optoelectronic devices to record and time-stamp insects' wingbeats. IEEE Sens J. 2016;16(15):6053–61.
84. Ouyang TH, Yang EC, Jiang JA, Lin T Te. Mosquito vector monitoring system based on optical wingbeat classification. Comput Electron Agric. 2015;118:47–55.
85. Moore A, Miller JR, Tabashnik BE, Gage SH. Automated identification of flying insects by analysis of wingbeat frequencies. J Econ Entomol. 1986;79(6):1703–6.
86. van Roy J, De Baerdemaeker J, Saeys W, De Ketelaere B. Optical identification of bumblebee species: Effect of morphology on wingbeat frequency. Comput Electron Agric. 2014;109:94–100.
87. Schmid B, Zaugg S, Votier SC, Chapman JW, Boos M, Liechti F. Size matters in quantitative radar monitoring of animal migration: estimating monitored volume from wingbeat frequency. Ecography. 2019;42(5):931–41.
88. Levin SA. Encyclopedia of Biodiversity. Elsevier Science; 2013.
89. Tauc MJ, Fristrup KM, Repasky KS, Shaw JA. Field demonstration of a wing-beat modulation lidar for the 3D mapping of flying insects. OSA Contin. 2019;2(2):332.
90. Brydegaard M, Svanberg S. Photonic monitoring of atmospheric and aquatic fauna. Laser Photon Rev. 2018;12(12):1800135.
91. Genoud AP, Saha T, Williams GM, Thomas BP. Insect biomass density: measurement of seasonal and daily variations using an entomological optical sensor. Appl Phys B [Internet]. 2023;129(2):1–13. Available from: https://doi.org/10.1007/s00340-023-07973-5

92. Malmqvist E, Jansson S, Zhu S, Li W, Svanberg K, Svanberg S, et al. The bat–bird–bug battle: Daily flight activity of insects and their predators over a rice field revealed by high-resolution scheimpflug lidar. R Soc Open Sci. 2018;5(4).
93. Santos V, Costa-Vera C, Rivera-Parra P, Burneo S, Molina J, Encalada D, et al. Dual-Band Infrared Scheimpflug Lidar Reveals Insect Activity in a Tropical Cloud Forest. Applied Spectroscopy, Vol 77, Issue 6, pp 593-602 [Internet]. 2023 Jun 1 [cited 2023 Jun 29];77(6):593–602. Available from: https://opg.optica.org/abstract.cfm?uri=as-77-6-593
94. Qi Y, Cinar GT, Souza VMA, Batista GE, Wang Y, Principe JC. Effective insect recognition using a stacked autoencoder with maximum correntropy criterion. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE; 2015. p. 1–7.
95. Fanioudakis E, Geismar M, Potamitis I. Mosquito wingbeat analysis and classification using deep learning. European Signal Processing Conference. 2018 Nov 29;2018-September:2410–4.

Paper I: Scheimpflug lidar range profiling of bee activity patterns and spatial distribution

Rydhmer *et al. Animal Biotelemetry* (2022) 10:14
https://doi.org/10.1186/s40317-022-00285-z

Animal Biotelemetry

METHODOLOGY Open Access

Scheimpflug lidar range profiling of bee activity patterns and spatial distributions

Klas Rydhmer[1,2*], Jord Prangsma[2], Mikkel Brydegaard[2,3,4], Henrik G. Smith[5], Carsten Kirkeby[2,6], Inger Kappel Schmidt[1] and Birte Boelt[7]

Abstract

Background: Recent declines of honeybees and simplifications of wild bee communities, at least partly attributed to changes of agricultural landscapes, have worried both the public and the scientific community. To understand how wild and managed bees respond to landscape structure it is essential to investigate their spatial use of foraging habitats. However, such studies are challenging since the foraging behaviour of bees differs between species and can be highly dynamic. Consequently, the necessary data collection is laborious using conventional methods and there is a need for novel methods that allow for automated and continuous monitoring of bees. In this work, we deployed an entomological lidar in a homogenous white clover seed crop and profiled the activity of honeybees and other ambient insects in relation to a cluster of beehives.

Results: In total, 566,609 insect observations were recorded by the lidar. The total measured range distribution was separated into three groups, out of which two were centered around the beehives and considered to be honeybees, while the remaining group was considered to be wild insects. The validity of this model in separating honeybees from wild insects was verified by the average wing modulation frequency spectra in the dominating range interval for each group. The temporal variation in measured activity of the assumed honeybee observations was well correlated with honeybee activity indirectly estimated using hive scales as well as directly observed using transect counts.

Additional insight regarding the three-dimensional distribution of bees close to the hive was provided by alternating the beam between two heights, revealing a "funnel like" distribution around the beehives, widening with height.

Conclusions: We demonstrate how lidar can record very high numbers of insects during a short time period. In this work, a spatial model, derived from the detection limit of the lidar and two Gaussian distributions of honeybees centered around their hives was sufficient to reproduce the observations of honeybees and background insects. This methodology can in the future provide valuable new information on how external factors influence pollination services and foraging habitat selection and range of both managed bees and wild pollinators.

Keywords: Lidar, Remote sensing, Entomology, Landscape ecology, Pollination, Honeybees

Background, motivation and aim

The decline of insect numbers in recent years has worried both researchers and the public [1]. Pollinators and in particular bees and hoverflies provide essential services in terms of pollination of wild plants [2] and crops [3]. Honeybees provide a large part of the pollination of crops, but wild pollinators are also quantitatively important crop pollinators [4] and are essential for wild plant pollination [5]. Accordingly, recent declines

*Correspondence: kkgr@ign.ku.dk
[1] Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark
Full list of author information is available at the end of the article

of honeybees [6] and simplifications of wild bee communities [7], has caused considerable concern [8]. The decline of wild pollinators has been attributed to a multitude of factors, such as landscape simplification causing loss of foraging and nesting habitat, increased use of pesticides, spread of diseases and potentially also direct competition with managed pollinators [8, 9]. The decline of managed bees is instead mostly related to socio-economic factors, including lack of profitability of bee keeping [10], which may, however, be related to landscape structure [11].

To generate a mechanistic understanding of how both wild pollinators and honeybees respond to landscape change and to monitor the pollination services they provide, it is essential to investigate their spatial use of foraging habitats. Bees are central place foragers, that have to find food for their offspring in the vicinity of their nests [12, 13]. For wild bees, a major reason for their decline is thought to be a loss of a continuous forage supply across the season and sufficiently close to the nest [14]. However, since bee species differ in their foraging ranges, the consequences of landscape simplification may be species dependent [15–17]. Similarly, the benefit of managing honeybees may depend on the forage landscape surrounding hives [11], with consequences for the interest of bee keepers to manage hives for honey production. Finally, honeybees and wild pollinators may to a smaller or larger extent share flower resources,, suggesting that they may compete [18, 19]. The scope and consequence of competition may depend on their foraging ranges [20], for example whether or not wide-ranging species such as honeybees are able to outcompete less mobile species in simplified landscapes [21]. Knowledge about the use of foraging habitat and mobility of bees is, therefore, essential when designing mitigation measures to counteract ongoing pollinator declines, e.g., to safeguard crop pollination.

Although knowledge of habitat selection and foraging ranges of bees is essential, there is a lack of information on how it varies between species, landscape types and over time. The major reason for this is that studies of habitat use and foraging movements are challenging. For example, their foraging may show spatio-temporal dynamics [22–24] that can differ between species [25, 26], resulting in a requirement of extensive data to describe their use of foraging habitat. Conventional methods to determine habitat use, such as pan-traps, fail to produce fine time-resolved data and may result in bias because of bees being attracted to the traps [27]. Other methods, such as Pollard walks, require considerable resources and may produce data that are so scarce that they need to be pooled over space or time for analyses [28]. Hence, there is a need for methods that allow for time-continuous monitoring of bees and can accurately resolve different taxonomic groups.

Detection of insects with radar was demonstrated as early as 1949 [28] and entomological use of radar has since been considerably refined [29, 30]. It has in particular been applied to monitor large insects, such as moths and locust, migrating at heights of hundreds of meters. Using existing weather radar infrastructure, large amounts of data can be made accessible for radar entomology [31, 32]. The monitoring of foraging insects close over the ground is challenged by ground clutter noise but harmonic radar systems [33, 34], where a nonlinear diode is glued to the insects, can track individual insects at low altitudes [35]. However, the technology is limited to monitoring insects strong enough to carry the antenna and is unsuitable for monitoring large numbers of insects.

Inspired by progress in entomological radar and early entomological lidar [36, 37], lidar entomology has evolved [38] and overcomes many of the challenges for remote monitoring of insects near the ground. Lasers and the shorter wavelengths used in lidar allow for increased sensitivity and superior beam control in terms of collimation and side lobes. This makes it possible to use lidar in cluttered environments, e.g., embedded in forest vegetation [39], or just above ground in agricultural fields [40, 41]. In recent years, it has been used in several applications due to its capability of recording large number of observations in short time [40, 42, 43]. Lidars can provide sufficient statistics of insect activity within minutes and the retrieval of modulation properties provide some discrimination between groups, although not yet to species level [43]. Lidar instrumentation has earlier been used to monitor honeybees [44] but to date there are no studies attempting to capture the whole foraging range throughout the day.

To evaluate the feasibility to monitor honeybee activity separately from the activity of other insects, we set up an entomological Scheimpflug lidar [45] to monitor the honeybee activity in a pollinator-dependent crop, white clover for seed production (*Trifolium repens L.*). In addition to the lidar measurements, the activity of honeybees was measured using modified Pollard walks [46] and hive scales, measuring the weight of the hives over time. In this paper, we aim to show the ability to distinguish between honeybee and general insect activity using a spatial model rather than individual classification of each insect observation.

Materials and methods

An entomological kHz lidar was used to monitor the honeybee and insect activity in a 755 m transect over a white clover field for seed production in Denmark on 4. to 6. July 2017. The field contained 6 clusters with ca 20

beehives each for pollination services, as shown in Fig. 1. For ground truthing, Pollard walks and hive scales were used to monitor the honeybee activity. The measurements were carried out from 11:20 to 20:35 on 4 July, from 08:50 to 20:15 on 5 July and 09:00–15:00 on 6 July (local summertime).

Study site

The study site was a 300*1000 m white clover field located on the island of Lolland, Denmark (54°46′ 15.7″ N 11°36′ 25.5″ E). This site was selected for its flatness. The field was surrounded by hedges along the long sides and a small deciduous forest in the far end (Fig. 1). Within the field, there were two flower strips with a mix of lacy phacelia (*Phacelia tanacetifolia)* and buckwheat (*Fagopyrum esculentum*) to attract and support insects. The white clover crop was established with an even plant density resulting in 1331 flowerheads per m^2 (average of six samples 12.5 × 50 cm). At the time of the experiment, the white clover was in full bloom. On the eastern side, a wheel track ran along the field. The surrounding area contained agricultural fields and small forests.

Lidar instrumentation

The lidar instrument was purchased from Norsk Elektro Optikk AS, Norway. It resembles the ones earlier described in [40, 47, 48]. Briefly, in this study, the light from a 3 W 808 nm laser diode was expanded using a beam expander with 500 mm focal length and 102 mm aperture. The emitted light was focused on a neoprene covered termination board at 755 m and a tree at 1000 m distance (Fig. 1). The back-scattered light from insects entering the beam was collected by a Newtonian telescope with 200 mm aperture and 800 mm focal length. To reduce the amount of background light in the system, the collected light was filtered by a 3 nm wide bandpass filter. The filtered light was recorded by a 2048-pixel (14 × 200 µm pixel size) silicon line scan camera mounted according to the Scheimpflug principle at 45° angle. The optical instrumentation was mounted on a tripod and protected from weather by a 3 × 3 m tent. Power was supplied by a mains connection from a residential house at the field border.

The laser beam was aimed ca. 2 m west of the southeastern beehive cluster ca 180 m along the beam (Fig. 1). On 6 July, the height of the beam was alternated in height between the termination plate and a tree at 1000 m every 15 min to profile the activity at two heights. The beam height above ground was measured on site at 15 locations along the transect and varied from ca 0.5 m close to the lidar to 2.5 m at the highest point for the lower beam. These measurements where combined with open source terrain data available from the Danish elevation model [49] and a linear model was used to interpolate the beam's height above the terrain along the full transect.

Data processing

The lidar recorded 35 000, 16-bit exposures at 3.5 kHz into a file of 10 s duration. The laser was synchronously modulated with the 3.5 kHz sampling frequency such that every second exposure was taken with the laser turned off. Between each file, there is an average gap of ~ 1 s due to data transfer which yields an average temporal fill-factor of ~ 90%. As in previous work [48], the frames, recorded when the laser is off, are subtracted from the frames recorded with the laser turned on yielding synchronized lock-in detection. This detection scheme allows the lidar to record insect echoes during

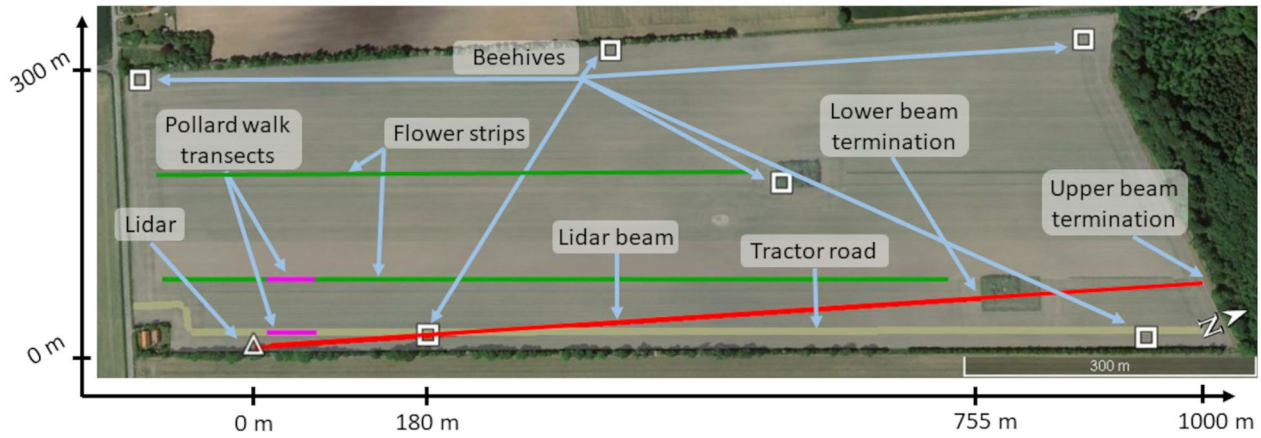


Fig. 1 Schematic of the field and experiment layout. The lidar beam passed ca 2 m from the beehives, ca 1.5 m above the ground. Satellite image from Google Earth

daytime by removing the influence of background illumination and yields a time-range map, as exemplified in Fig. 2. Subsequently, the 10 s median intensity at each pixel is subtracted to remove static signals from atmospheric backscattering.

The Scheimpflug ranging principle is based on triangulation and thus the pixel number corresponds tangentially to the range [50–54]. Insect observations, as the one shown in Fig. 2b, were automatically extracted from the raw data using a slightly adapted version of the algorithm described in detail in [48]. An insect observation is defined as a sequence of above-threshold signals produced when an insect transits the beam. Each observation thus consists of a single insect trajectory through the beam. In total over 3 days, 566 609 individual insect observations were recorded by the lidar during a total measurement time of 23 h and 15 min.

Ground truthing

Two 45 m modified Pollard walks were conducted every hour at two different transects ca 150 m from the beehives, as shown in Fig. 1 [46]. The first was a 100 cm wide area between two wheel tracks for field operations in the white clover. The second was a 150 cm wide flower strip with phacelia and buckwheat. The transect walks were conducted by various operators at the site, in total, 5 different persons, counting all honey- and bumblebees foraging, resting or flying between the wheel tracks or within the flower strip. Due to a limited number of bumblebees observed (< 50 in total), we only used honeybee counts in the further analyses. On average, ca 40 bees were observed by each observer and Pollard walk and in total, 5730 honeybee observations were made.

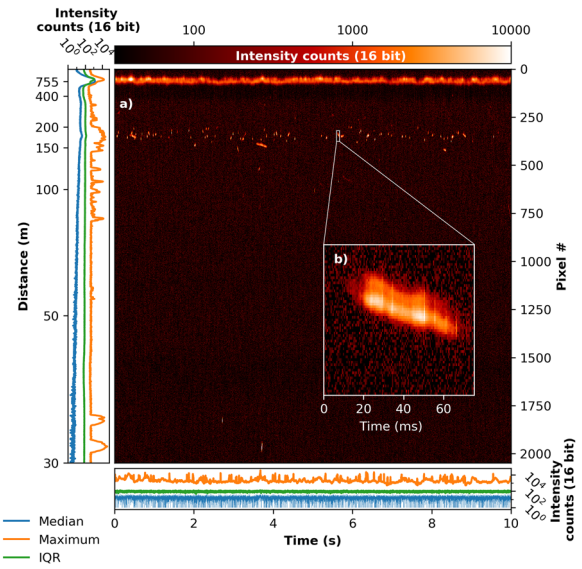


Fig. 2 **a** Lidar raw data example. The time-range map reveals a large number of insect observations by the beehives at 180 m. At the top of the image, the static echo from the termination is visible as a continuous line at 755 m. The signal intensity over range is shown in the vertical plot to the left with maximum, median and interquartile range (IQR) signals in orange, blue and green. Likewise, the signal over time (up to 700 m) is shown in the horizontal plot at the bottom. **b** Cut out showing an insect observation, where the wingbeats are visible as vertical stripes

Two of the beehives were remotely monitored by the beekeeper and the weight was logged every second hour. The change in weight over the course of the day is determined by the number of bees in the hive as well as the amount of collected pollen and nectar. We assume that all bees are inside the hives at midnight, thus representing the total weight of bees and the hive, W_{tot}. By linearly interpolating the change in hive weight from midnight to midnight, we can then, for every two hours, subtract the measured "hive weight", W_{Hive}, which is removed from each weight measurement:

$$W_{\mathrm{Bee}} = W_{\mathrm{tot}} - W_{\mathrm{Hive}} \quad (1)$$

where W_{Bee} represents the lost weight of the bees in the hive when they are out foraging, assumed to be directly correlated to the number of bees in the hive and, therefore, negatively correlated to the flight activity. This weight loss will of course also be affected by the amount of pollen and nectar collected and consumed between each 2-h sample, this is ignored in our model.

Weather data was collected by a small weather station with 30-min resolution monitoring temperature, humidity, air pressure, wind speed and wind direction. In general, the weather was stable with temperatures between 15 and 25 degrees Celsius, varying sun and cloud coverage and low winds during the entire measurement period.

Measurement results and data analysis

The distribution of insect recordings over time and range is shown in Fig. 3. Half of the observations were recorded within 11 m of the beehives. The maximum activity recorded by the lidar was reached between 14:45 and 15:00 with a total of 10 807 insect observations along the whole transect and 26 insect observations per meter and minute.

The measured insect activity over range during the peak activity is plotted in Fig. 4. From the gathered lidar data, we hypothesized that the spatial distribution of insect activity can be explained by three types of observations: hive activity, due to honeybees flying around near the hives, honeybees foraging within the field and background activity from wild insects. The activity from male drones is neglected in this model, since they generally only make up less than 10% of the total population in a beehive [55, 56]. Drones can aggregate in drone

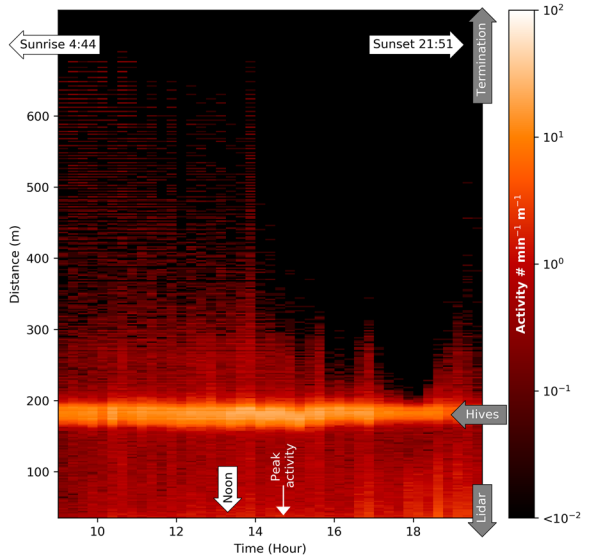


Fig. 3 Time-range map of insect activity during the second measurement day. Insect counts per minute and 1 m of transect evaluated in 15 min, 2 m bins. By the beehives at 180 m, the activity is 30 times higher than in the surrounding area. More activity is recorded closer to the lidar, since the sensitivity decreases with range. The maximum activity is indicated by an arrow at 14:45

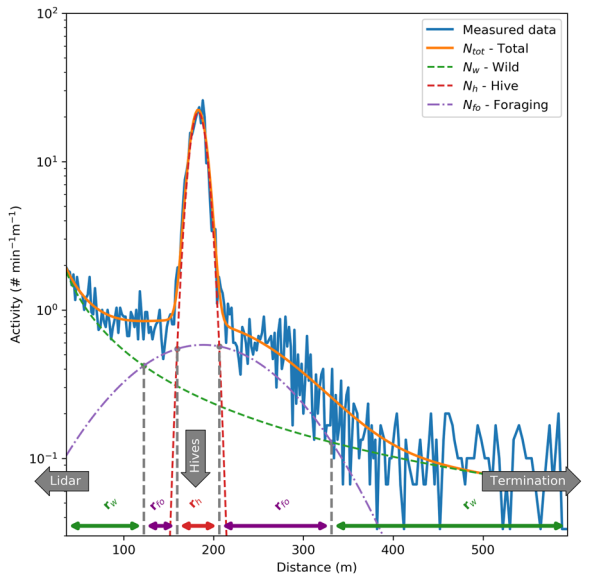


Fig. 4 Range distribution of detected insect observations between 14:45 and 15:00, 5/7 2017. The range distribution is approximated by three distributions. The peak in activity next to the beehives at 180 m is fitted by the red curve N_h and assumed to be the result of bees clustering near the hives. Foraging bees are fitted by the purple curve N_{fo}. The distribution from background insects not centered around the beehives, N_w, is plotted in green

congregation areas (DCAs), but these generally occurs at higher altitudes than we've monitored in this study [57]. To quantify the honeybee and wild insect activity we used a spatial model to decompose the observed range distribution into these three components. In simple terms: insect distributions centered around the beehives are assumed to be either clustering, or foraging bees. This is modelled as

$$(r,t) = N_w(r,t) + N_{fo}(r,t) + N_{\text{hive}}(r,t) \tag{2}$$

where N_w is the number of wild insects, N_{fo} is the number of foraging honeybees and N_{hive} is the hive activity from honeybees located near the hives.

The distribution of wild insects is defined as a negative exponential function:

$$N_w(r,t) = N_{0w}(t)r^{\alpha(t)}, 35 < r < 755 \tag{3}$$

where r is the range from the lidar, and α is a negative parameter which depends on the optical properties of the targets. This reciprocal distribution is caused by the reduced sensitivity of the lidar with range and the expected measured result from insects distributed homogenously in the field [54].

The hive activity and foraging honeybees are modelled as Gaussian distributions, N_h and N_{fo}, centered around the beehive cluster:

$$N_{fo}(r,t) = N_{0fo}(t)e^{-\frac{(r-r_c(t))^2}{R(t)^2}} \tag{4}$$

where N_{0fo} is the maximum number of observations, r_c is the centre position and R is the width of the curve.

The model in Eqs. (2)–(4) has 8 free parameters and was fitted to aggregated range distributions with a bin width of 2 m, yielding 360 datapoints from 35 to 755 m. using Scipy's optimization package [58]. The model was fitted to 15 min subsets of the collected insect observations and had an average adjusted r-squared correlation coefficient of 0.96 with a standard deviation of 0.026.

The foraging distribution (shown in purple in Fig. 4) includes both flights to and from foraging sites as well as actual foraging flights. The width of the foraging distribution describes the foraging range from the beehives and has an average full width half maximum of 153 m throughout the full measurement period, with a standard deviation of 50 m.

All insect observations were split into three groups, matching the regions, where N_w, N_h or N_{fo} dominated the model as illustrated in Fig. 4. To investigate the assumption that these groups consist of different insect species, we estimated the modulation powers of a sample of insects selected in the dominating range interval of each group by the Welch method [59]. In Fig. 5, the median

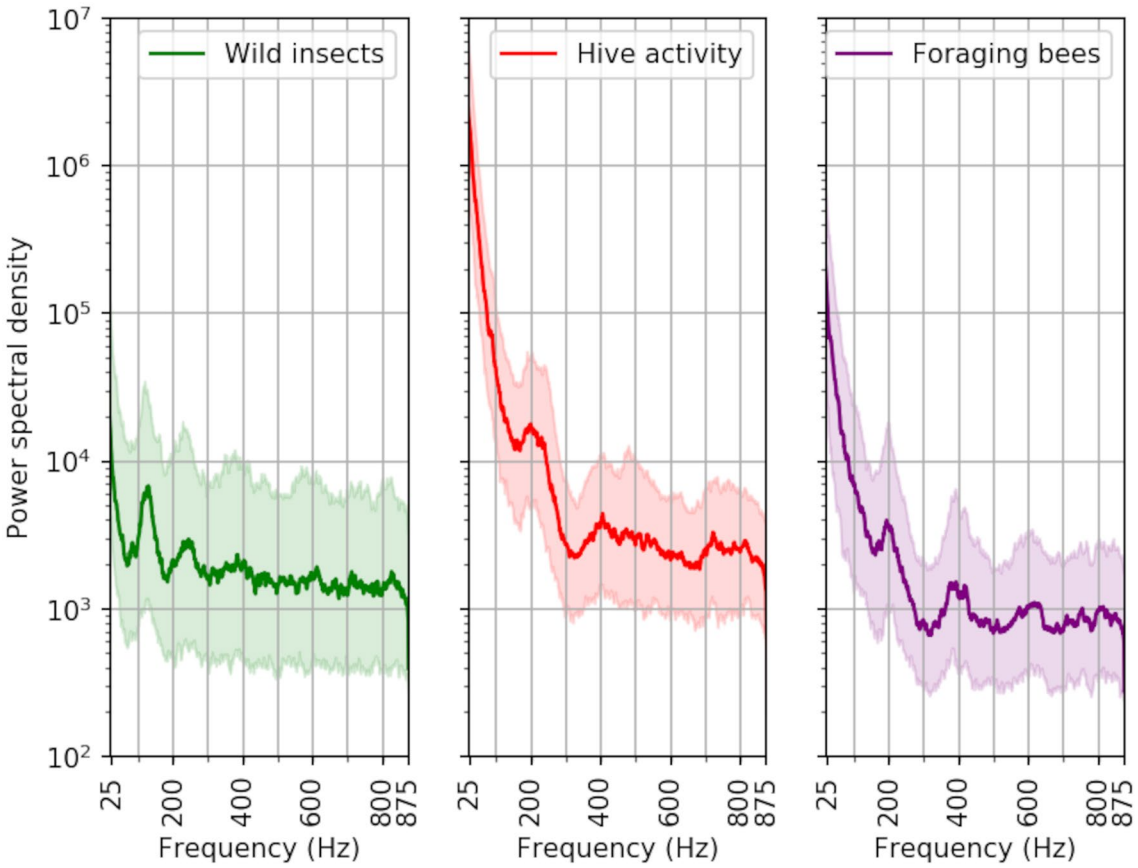


Fig. 5 Modulation power frequency distributions. Green curve represents the wild insects observed in r_w region, red curve represents the hive activity in the r_h region. The purple curve represents the foraging bees observed in the r_{fo} region. Median intensity is plotted as a solid line and the IQR between 25 and 75% as a band. The insect observations recorded in the r_h and r_{fo} regions are assumed to be honeybees in the model. This is supported by the spectra, since they show a strong peak between 180 and 220 Hz

power spectra from 500 random observations recorded in r_w, r_h and r_{fo} is presented. We see that both honeybee distributions have a strong peak around 180–220 Hz which fits well with the expected wingbeat frequency for honeybees from literature [60–62]. Wild insects show a different distribution with lower and more varied wingbeat frequencies than the clustered and foraging bees. We only counted honeybees and bumble bees during the Pollard walks but many other small insects can be expected to be active in the field.

Fitting the spatial model to all data, the number of wild insects, clustering and foraging bees can be estimated for the full measurement period. The result is plotted together with the ground truthing results in Fig. 6. The lidar was shut down for ca 20 min due to computer problems on 4 July around 11:00 and thus, some data points are missing in Fig. 6. The lidar data from 6 July is shown separately in Fig. 7.

In Fig. 6a, b, we see that the hive activity and foraging bees show a strong daily pattern, where activity rises during the morning, reaches peak activity around 14:00 both days, slightly after the solar noon at 13:18 [63] and decreases in the afternoon. In contrast, the wild insect activity shows a more consistent activity throughout the day, and even increases throughout the entire second day.

The lidar measurements of bee activity show good correlation with the reference measurements (Table 1). The correlation between the Pollard walk counts and hive scale measurements were calculated by linearly interpolating between the two closest sample points of the hive scales to each Pollard walk. The correlation between the Pollard walks and the lidar was calculated by interpolating the between the two closest 15 min recording intervals to each Pollard walk. Since the lidar was alternated between a higher and lower transect during the third day, it gradually became un-aligned and recorded fewer and fewer observations in each timeslot. Lidar data from 6 July is, therefore, not comparable to the two previous days or used in the correlation calculations.

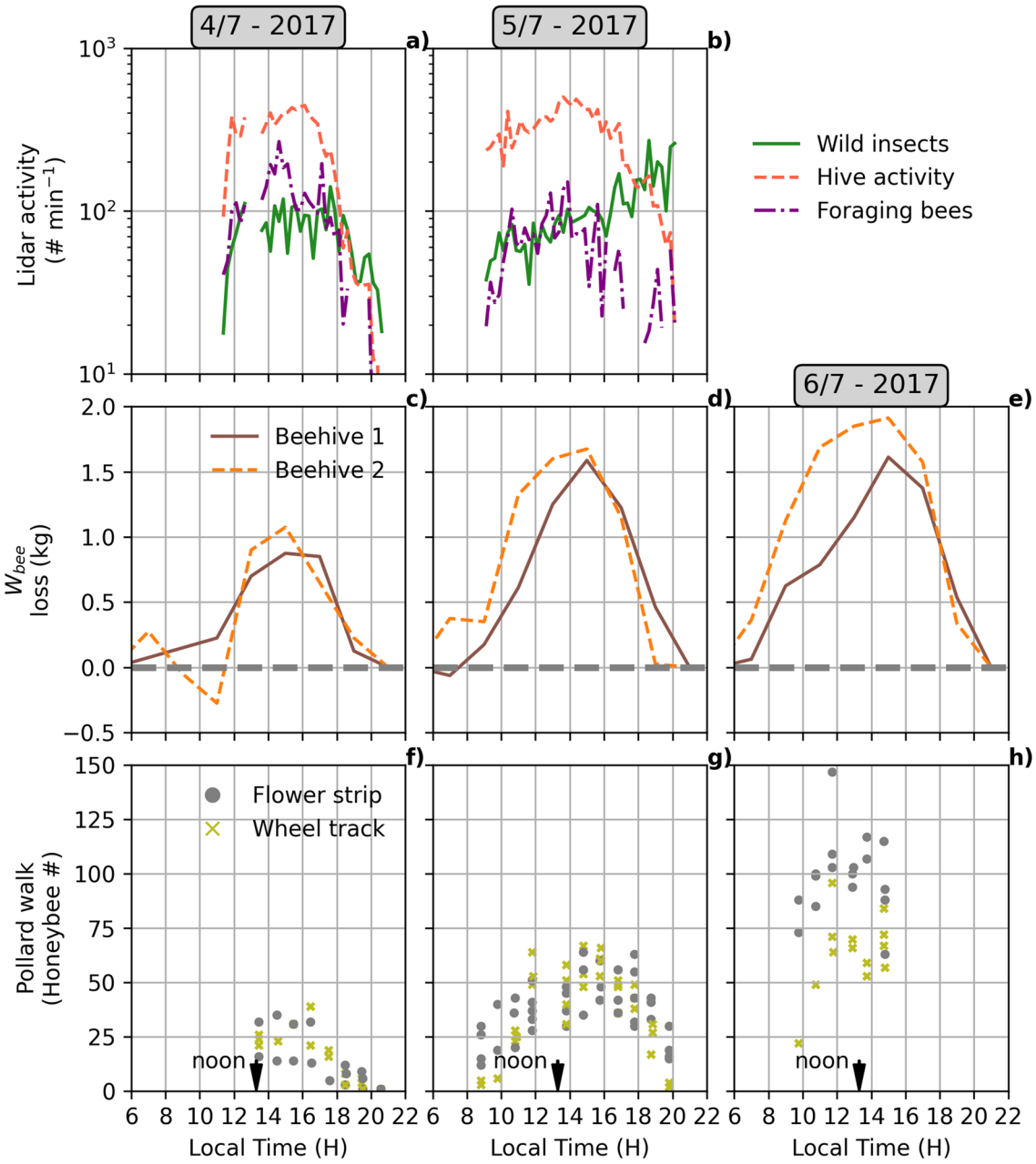


Fig. 6 Lidar insect counts, W_{bee} loss, and Pollard walk observations over time. The clustered and foraging bees detected by the lidar show a different temporal pattern than the background insects, which are active later in the evening. This correlates with the estimate of bee activity determined from hive scales and with the transect counts

The transect walks by the wheel track show a slightly better correlation with hive and lidar measurements than the flower path, possibly because the honeybees were easier to spot in the low clover crop than in the flower strip. The lidar shows slightly better correlation with the loss in hive weights than with the Pollard walk observations. This is also shown in Fig. 8.

Analyzing lidar data from the third day when the beam was altered between two transects, we find ~ 68% more bees in the lower transect than in the upper. The bees observed near the beehives in the upper transect are also more dispersed, as shown in Fig. 7a, b. The average wingbeat frequency spectra in the lower and, respectively, upper transect are shown in Fig. 7c. By fitting a spectral

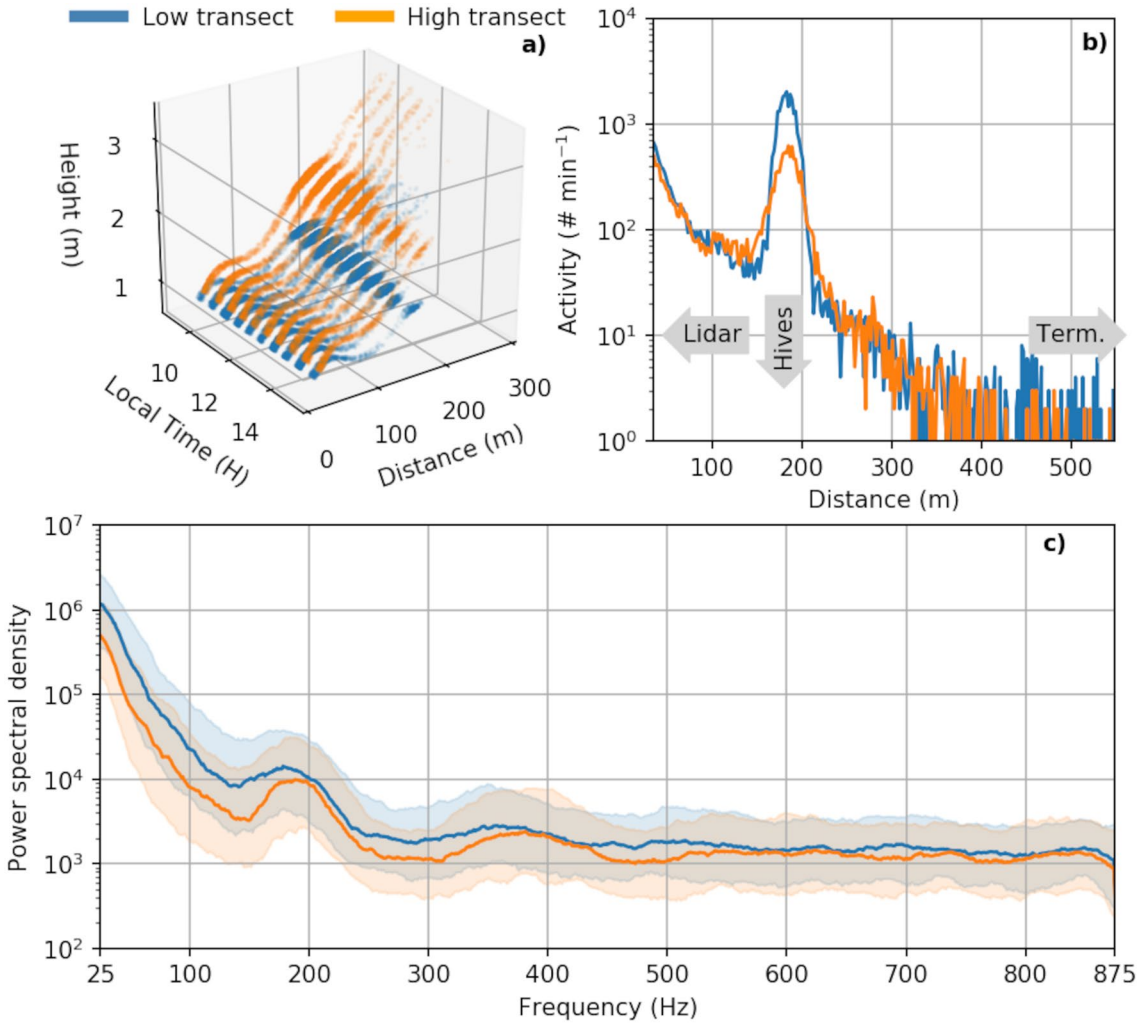


Fig. 7 **a** Insect activity over time, range and flight height. Each dot represents one observation. Observations recorded in the lower transect are plotted in blue. Observations from the higher transect are plotted in orange. The majority of observations are recorded close to the beehive at 180 m. Note that the variation in ground level is exaggerated by the different scales on the height and distance axes. **b** Distribution over range during the third day for both transects. In the lower transect, we see a high number of bees concentrated in a small area, whereas in the higher transect, we see fewer and a broader distribution of insects which indicates a "funnel like" distribution above the beehives. **c** Median frequency composition and IQR for the low and high transect within 15 m of the beehives. Insect in the high transect have a 7% higher average wingbeat frequency. The lower transect also shows more activity at lower frequencies from wild insects

model from [64], the fundamental frequency in the lower and, respectively, upper transect can be estimated. With an explanation grade for the model of > 98%, the fundamental wingbeat frequency was calculated to 179.6 Hz with in the lower transect and 191.5 Hz in the upper transect. Confidence intervals were 178.8 Hz to 180.4 Hz and 191 Hz to 192 Hz, respectively.

Discussion

In this study, we have separated hive activity and foraging bees from wild insects. The measured activity is correlated with alternative measures of activity obtained by hive scales and Pollard walks. The average foraging distance is estimated and the insect distribution close to the hives is profiled by multiplexing the height of the beam.

The large abundance of honeybees in this experimental setup made it possible to assume that the vast majority of insects centered around the beehives were due to hive activity or foraging honeybees. While there were several beehive clusters in the field, the selected cluster was relatively isolated from the others and we could assume that insects showing a different spatial distribution were other insects. This made it possible to calculate the honeybee activity and foraging range without individually

Table 1 Relationships (Pearson correlation coefficient R, *p* value p and number of time intervals used N) between alternative measures of honeybee activity

	Pollard walk counts, wheel track	**Pollard walk counts, flower path**	**W_{bee} loss hive 1**	**W_{bee} loss hive 2**
Pollard walk counts, flower path	R: 0.811 p: 5.1×10^{-7} N: 26			
W_{bee} loss hive 1	R: 0.846 p: 1.08×10^{-9} N: 32	R: 0.638 p: 8.43×10^{-5} N: 32		
W_{bee} loss hive 2	R: 0.874 p: 6.76×10^{-11} N: 32	R: 0.832 p: 3.69×10^{-9} N: 32	R: 0.821 p: 2.30×10^{-10} N: 38	
Lidar bee counts ($N_{hive} + N_{fo}$)	R: 0.607 p: 1.65×10^{-3} N: 24	R: 0.526 p: 8.29×10^{-3} N: 24	R: 0.725 p: 8.61×10^{-6} N: 29	R: 0.777 p: 7.00×10^{-7} N: 29

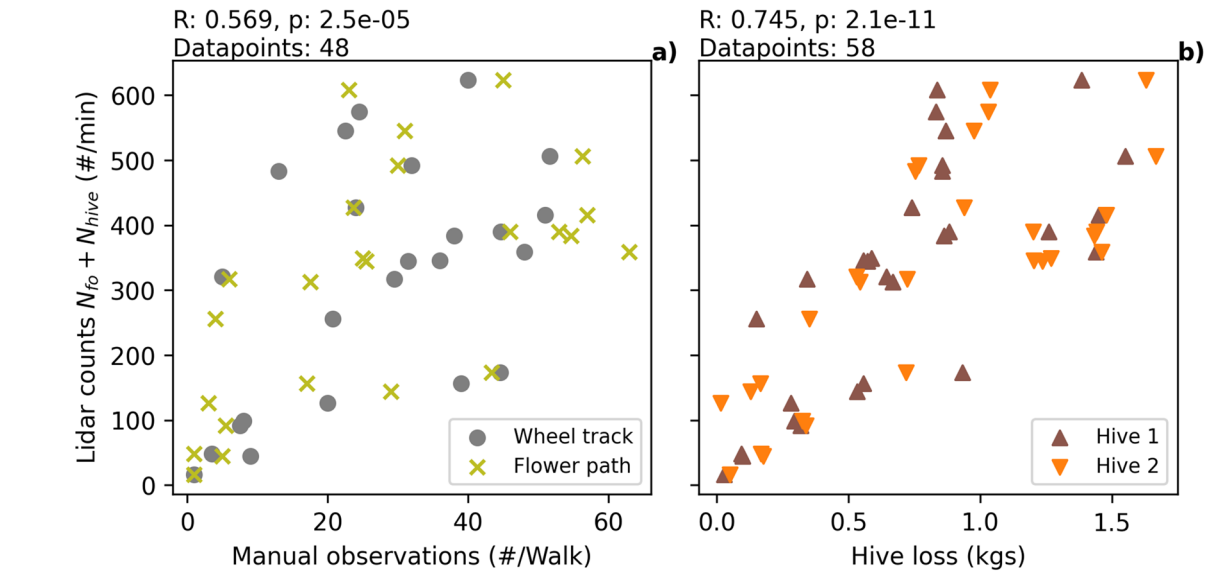


Fig. 8 Measured honeybee activity from the lidar is correlated with the observations from the transect walks (**a**). The correlation is even stronger between the lidar and the hive weights (**b**). The interpretation could be that the lidar and hive scales measure all flight activity, whereas the Pollard walks only measure bees foraging in part of the field

classifying each observation. This simplification was validated by the frequency spectra shown in Fig. 5. The drone activity was ignored as they only make up a relatively small fraction of the individuals in a beehive.

The lidar, beehive scales and manual transect walks all show good agreement on the honeybee activity (Table 1). However, using the beehive scales to monitor activity is based on the approximation that the weight is linearly changing by a constant rate from midnight to midnight. This is an assumption; the weight of the hives depends on the feed brought into the hive, the feed eaten and the weight of the bee population within the hive. The lidar measurements are more strongly correlated with the hive scales than the Pollard transects. One interpretation is that the hive scales and lidar measure all flight activity, while the Pollard transects mainly record flower visits and flights close to the ground level. Alternatively, the Pollard counts are prone to more random variation, caused by observations of shorter time duration, smaller spatial scale covered and bee detectability. In addition, both the lidar and the hive scales are more strongly correlated with the activity in the wheel track than in the flower strip. This could indicate that the bees from the monitored

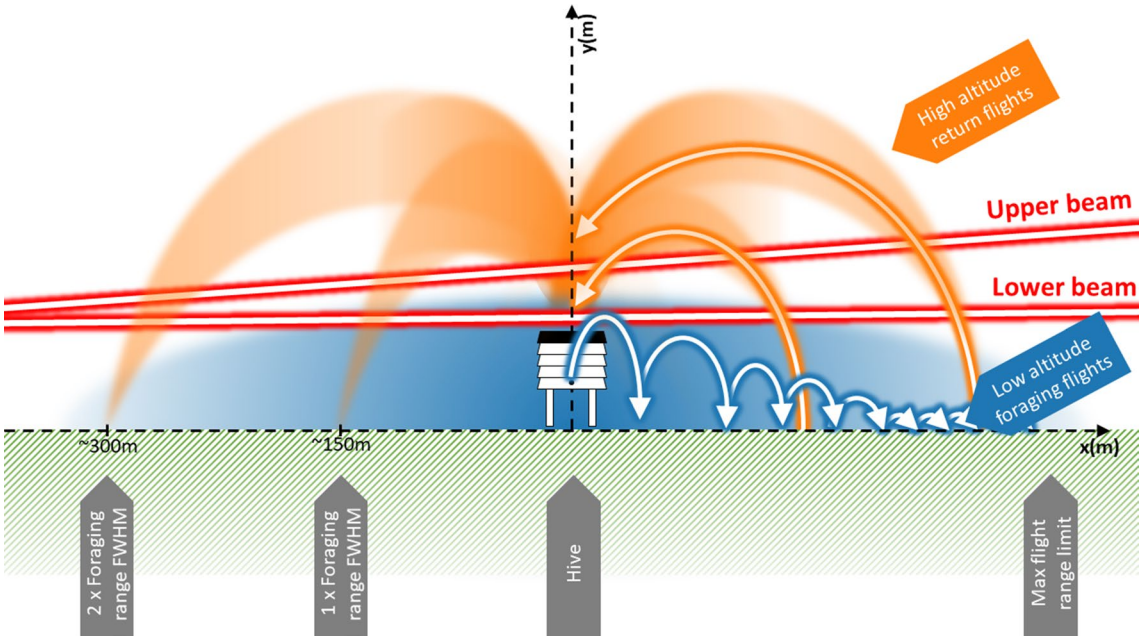


Fig. 9 Funnel-like distribution and differences in wingbeat frequency could be explained by a model, where insects leaving the hive fly at lower height, close to the ground, while returning bees fly along a higher path. In this study, the full-width-half-maximum (FWHM) of the foraging range was ca 150 m

hive were mainly foraging in the nearby white clover, or that the honeybees were more difficult to count in the high growing flower strip.

By modulating the beam between a lower and higher path, we profiled the honeybee activity at two heights. The fewer and more dispersed insects in the upper beam seems to indicate a "funnel like" distribution of bees over the beehives, widening with height, as shown in Fig. 7a, b. Orientation flights of new workers have been described as a spiral widening with height and could contribute to this distribution [65]. However, the number of orientation flights is expected to be low compared to the number of foraging flights. In addition, the insects observed in the upper beam had 7% higher wingbeat frequency, as can be seen in Fig. 7c. If returning bees carrying nectar and pollen have a higher wingbeat frequency, the results indicate a scenario, where bees leave the hives flying close to the ground for foraging. Once fully loaded, ca 150 m from the hive, they return to the hive on a higher trajectory, as illustrated in Fig. 9. However, a recent study based on a limited number of measurements failed to find a correlation between weight load and wingbeat frequency [61]. Other work on bumble bees show that payload initially affect flight pitch angle and that the wing beat frequency is only increased in extreme cases [66]. Wingbeat frequency increases with temperature, and since the air is expected to be warmer near the ground, a thermal difference between the beams is unlikely to be the cause.

A few previous experiments used a "scanning" beam to map insect activity in three dimensional space [67], but to our knowledge this is the first time this is combined with automated algorithms for individual event extraction on a large number of observations. In this work, the beam was moved manually and only vertically but the logical progression would be to alternate the beam horizontally over more transects and automate the movement. One could also employ a 2D detector chip in combination with a laser sheet [68]. This would allow a 2D model of the foraging range within the field. We wish to explore this in future studies. Although this experiment only covered total of ~ 23 h of recordings, the high number of collected recordings allows statistical analysis of temporal changes with 15 min resolution. This is to the best of our knowledge not possible with any other insect monitoring method.

In this study, the range distribution of the foraging bees had an average full width half maximum of ~ 150 m. This can be compared to studies in the literature which finds that foraging ranges vary from 45 to 6000 m, with average foraging distances typically around 600 m to 800 m depending on colony size, foraging resources and time of year, with shorter distances in early summer [69, 70]. In this field, the hives were placed in the middle of a food source which has been shown to result in shorter foraging distances [23]. However, as discussed in Fig. 4, the sensitivity of the lidar decreases with range due to the

optical configuration. Therefore, the minimum detectable target size decreases with range, and in addition, the beam's elevation over the crop is also varying along the transect. This makes it hard to quantitatively compare the activity at different distances [23]. A future more complex parameterization model could take these parameters into account as discussed in [54] and investigate inhomogeneous distributions of wild insects, pollinator competition and displacement of wild pollinators.

Outlook

Since the height of the beam strongly influences the number of detected observations, it is challenging for lidar entomologists to compare insect activity levels at different locations. Regardless, the instrumentation can be a vital tool to investigate the behavior of bees and wild insects. In this study, a simple spatial model was relied on to discriminate target types and provide quantitative estimates of their relative occurrence. As characterization of the scattering properties of individual insects develops, discrimination at the level of individual transit observations may become possible [71–73].

While instrumentation used in this study is commercially available, it currently requires skilled technicians for alignment and operations. As the entomological lidar community is growing and research groups are active in several countries and continents, there are good prospects for the methodology becoming accessible for entomologists and ecologists in general.

Conclusions

We deployed an entomological lidar in a homogenous flowering white clover field and profiled the honeybee activity around a cluster of beehives over time. By decomposing the observations into hive activity, foraging honeybees and wild insects the number of honeybees engaged in flight activities could be estimated and showed good correlation with estimates from hive scales and Pollard walks. In addition to counting the number of active bees, average foraging distance was estimated. In addition, the three-dimensional distribution of honeybees around the hives was investigated by moving the beam between an upper and lower height.

This work has shown the ability to record very high number of insects during a short time period, which allows the study of insect activity with a very high temporal resolution. We propose that lidar monitoring can change pollinator research in the future by providing valuable new information on how external factors influence pollinator activity.

Acknowledgements
We thank Frederik Taarnhøj for assisting with the grant proposal, Flemming Rasmussen, Per Kryger, Janne Kool, Alem Gebru and Josephine Nielsen for assistance during the field measurements. We thank Krenkerup Estate for access to the white clover seed crop and Mr. Søren Jespersen and Mathias Knudsen for information on crop management and support setting up the equipment in the field.

Author contributions
KR, JP and BB conceived the experiment and acquired the grant. MB constructed the lidar instrument and wrote the initial analysis code. KR and JP carried out the experiment on site. KR and HS drafted the manuscript. CK, MB and IS directed data analysis. KR analyzed data and produced graphical items. All authors read and approved the final manuscript.

Funding
This study was supported by a grant by Idagaardfonden, Denmark, Innovation Foundation, Denmark, the Swedish Research Council, Norsk Elektro Optikk AS, Norway, Formas Sweden and 15. Juni and Aage V. Jensen Nature Foundations, Denmark.

Availability of data and materials
The data set used will be made available in a reduced format at a public repository at publication. In addition, contact information to potential regional entomological research groups can be acquired from the authors by request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
KR, CK, JP and MB are presently or formerly affiliated, employees or shareholders of FaunaPhotonics. We declare that this has not affected the reported results or interpretations in any way.

Author details
[1]Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark. [2]FaunaPhotonics APS, Støberigade 14, 2450 Copenhagen, SV, Denmark. [3]Lund Laser Centre, Department of Physics, Lund University, Sölvegatan 14, 223 62 Lund, Sweden. [4]Norsk Elektro Optikk AS, Prost Stabels vei 22, N-2019 Skedsmokorset, Norway. [5]Centre of Environmental and Climate Science & Department of Biology, Lund University, Sölvegatan 35, 223 62 Lund, Sweden. [6]Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 8, 1870 Frederiksberg, Denmark. [7]Department of Agroecology - Crop Health, Aarhus University, Forsøgsvej 1, building 7610, A132, 4200 Slagelse, Denmark.

Received: 9 April 2021 Accepted: 1 April 2022
Published online: 19 April 2022

References

1. Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, et al. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE. 2017;12:10.
2. Ollerton J. Pollinator diversity: distribution, ecological function, and conservation. Annu Rev Ecol Evol Syst. 2017;48:353–76.
3. Klein A-M, Vaissiere BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, et al. Importance of pollinators in changing landscapes for world crops. Proc R Soc B Biol Sci. 2007;274(1608):303–13.
4. Garibaldi LA, Steffan-Dewenter I, Winfree R, Aizen MA, Bommarco R, Cunningham SA, et al. Wild Pollinators Enhance Fruit Set of Crops Regardless of Honey Bee Abundance. Science. 2013 Mar 29;339(6127):1608 LP – 1611. http://science.sciencemag.org/content/339/6127/1608.abstract

5. Stanley DA, Msweli SM, Johnson SD. Native honeybees as flower visitors and pollinators in wild plant communities in a biodiversity hotspot. Ecosphere. 2020;11:2.
6. Potts SG, Roberts SPM, Dean R, Marris G, Brown MA, Jones R, et al. Declines of managed honey bees and beekeepers in Europe. J Apic Res. 2010;49(1):15–22.
7. Powney GD, Carvell C, Edwards M, Morris RKA, Roy HE, Woodcock BA, et al. Widespread losses of pollinating insects in Britain. Nat Commun [Internet]. 2019;10(1):1–6. https://doi.org/10.1038/s41467-019-08974-9.
8. Potts SG, Ngo HT, Biesmeijer JC, Breeze TD, Dicks L V, Garibaldi LA, et al. The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. 2016;
9. Goulson D, Nicholls E, Botías C, Rotheray EL. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science. 2015;347(6229):1255957.
10. Breeze TD, Boreux V, Cole L, Dicks L, Klein A, Pufal G, et al. Linking farmer and beekeeper preferences with ecological knowledge to improve crop pollination. People Nat. 2019;1(4):562–72.
11. Sponsler DB, Johnson RM. Honey bee success predicted by landscape composition in Ohio USA. PeerJ. 2015;3:e838.
12. Kacelnik A, Houston AI, Schmid-Hempel P. Central-place foraging in honey bees: the effect of travel time and nectar flow on crop filling. Behav Ecol Sociobiol. 1986;19(1):19–24.
13. Cresswell JE, Osborne JL, Goulson D. An economic model of the limits to foraging range in central place foragers with numerical solutions for bumblebees. Ecol Entomol. 2000;25(3):249–55.
14. Smith HG, Birkhofer K, Clough Y, Ekroos J, Olsson O, Rundlöf M. Beyond dispersal: the role of animal movement in modern agricultural landscapes. In: Animal Movement Across Scales. Oxford University Press; 2014. p. 51–70.
15. Steffan-Dewenter I, Münzenberg U, Bürger C, Thies C, Tscharntke T. Scale-dependent effects of landscape context on three pollinator guilds. Ecology. 2002;83(5):1421–32.
16. Persson AS, Rundlöf M, Clough Y, Smith HG. Bumble bees show trait-dependent vulnerability to landscape simplification. Biodivers Conserv. 2015;24(14):3469–89.
17. Bommarco R, Lundin O, Smith HG, Rundlöf M. Drastic historic shifts in bumble-bee community composition in Sweden. Proc R Soc B Biol Sci. 2012;279(1727):309–15.
18. Herbertsson L, Lindström SAM, Rundlöf M, Bommarco R, Smith HG. Competition between managed honeybees and wild bumblebees depends on landscape context. Basic Appl Ecol. 2016;17(7):609–16.
19. Thomson DM, Page ML. The importance of competition between insect pollinators in the Anthropocene. Curr Opin Insect Sci. 2020;38:55–62.
20. Westphal C, Steffan-Dewenter I, Tscharntke T. Bumblebees experience landscapes at different spatial scales: possible implications for coexistence. Oecologia. 2006;149(2):289–300.
21. Bolin A, Smith HG, Lonsdorf EV, Olsson O. Scale-dependent foraging tradeoff allows competitive coexistence. Oikos. 2018;127(11):1575–85.
22. Herrera CM. Daily patterns of pollinator activity, differential pollinating effectiveness, and floral resource availability, in a summer-flowering Mediterranean shrub. Oikos. 1990;1:277–88.
23. Danner N, Molitor AM, Schiele S, Härtel S, Steffan-Dewenter I. Season and landscape composition affect pollen foraging distances and habitat use of honey bees. Ecol Appl. 2016;26(6):1920–9.
24. Pope NS, Jha S. Seasonal food scarcity prompts long-distance foraging by a wild social bee. Am Nat. 2018;191(1):45–57.
25. Willmer PG, Bataw AAM, Hughes JP. The superiority of bumblebees to honeybees as pollinators: insect visits to raspberry flowers. Ecol Entomol. 1994;19(3):271–84.
26. Redhead JW, Dreier S, Bourke AFG, Heard MS, Jordan WC, Sumner S, et al. Effects of habitat composition and landscape structure on worker foraging distances of five bumble bee species. Ecol Appl. 2016;26(3):726–39.
27. Baum KA, Wallen KE. Potential bias in pan trapping as a function of floral abundance. J Kansas Entomol Soc. 2011;84(2):155–9.
28. Garratt MPD, Senapathi D, Coston DJ, Mortimer SR, Potts SG. The benefits of hedgerows for pollinators and natural enemies depends on hedge quality and landscape context. Agric Ecosyst Environ. 2017;247:363–70.
29. Drake VA, Reynolds DR. Radar entomology: observing insect flight and migration. Cabi; 2012.
30. Daniel Kissling W, Pattemore DE, Hagen M. Challenges and prospects in the telemetry of insects. Biol Rev. 2014;89(3):511–30.
31. Westbrook JK, Eyster RS, Wolf WW. WSR-88D doppler radar detection of corn earworm moth migration. Int J Biometeorol. 2014;58(5):931–40.
32. Gauthreaux SA Jr, Livingston JW, Belser CG. Detection and discrimination of fauna in the aerosphere using Doppler weather surveillance radar. Integr Comp Biol. 2008;48(1):12–23.
33. Riley JR, Valeur P, Smith AD, Reynolds DR, Poppy GM, Löfstedt C. Harmonic radar as a means of tracking the pheromone-finding and pheromone-following flight of male moths. J Insect Behav. 1998;11(2):287–96.
34. Ovaskainen O, Smith AD, Osborne JL, Reynolds DR, Carreck NL, Martin AP, et al. Tracking butterfly movements with harmonic radar reveals an effect of population age on movement distance. Proc Natl Acad Sci. 2008;105(49):19090–5.
35. Riley JR, Smith AD, Reynolds DR, Edwards AS, Osborne JL, Williams IH, et al. Tracking bees with harmonic radar. Nature. 1996;379(6560):29–30. https://doi.org/10.1038/379029b0.
36. Shaw JA, Seldomridge NL, Dunkle DL, Nugent PW, Spangler LH, Bromenshenk JJ, et al. Polarization lidar measurements of honey bees in flight for locating land mines. Opt Express. 2005;13(15):5853–63.
37. Guan Z, Brydegaard M, Lundin P, Wellenreuther M, Runemark A, Svensson EI, et al. Insect monitoring with fluorescence lidar techniques: field experiments. Appl Opt. 2010;49(27):5133–42.
38. Brydegaard M, Svanberg S. Photonic monitoring of atmospheric and aquatic fauna. Laser Photon Rev. 2018;12(12):1800135.
39. Li M, Jansson S, Runemark A, Peterson J, Kirkeby CT, Jönsson AM, et al. Bark beetles as lidar targets and prospects of photonic surveillance. J Biophotonics. 2020;1:1–16.
40. Brydegaard M, Gebru A, Kirkeby C, Åkesson S, Smith H. Daily evolution of the insect biomass spectrum in an agricultural landscape accessed with lidar. In: EPJ Web of Conferences. EDP Sciences; 2016. p. 22004.
41. Malmqvist E, Jansson S, Zhu S, Li W, Svanberg K, Svanberg S, et al. The bat–bird–bug battle: Daily flight activity of insects and their predators over a rice field revealed by high-resolution scheimpflug lidar. R Soc Open Sci. 2018;5:4.
42. Brydegaard M, Jansson S. Advances in entomological laser radar. IET Int Radar Conf. 2018;(Irc 2018):2–5.
43. Brydegaard M, Jansson S, Malmqvist E, Mlacha Y, Gebru A, Okumu F, et al. Lidar reveals activity anomaly of malaria vectors during pan-African eclipse. Sci Adv. 2020;13:6.
44. Hoffman DS, Nehrir AR, Repasky KS, Shaw JA, Carlsten JL. Range-resolved optical detection of honeybees by use of wing-beat modulation of scattered light for locating land mines. Appl Opt. 2007;46(15):3007–12.
45. Brydegaard M, Malmqvist E, Jansson S, Larsson J, Török S, Zhao G. The Scheimpflug lidar method. In: Lidar Remote Sensing for Environmental Monitoring 2017. International Society for Optics and Photonics; 2017. p. 104060I.
46. Pollard E. A method for assessing changes in the abundance of butterflies. Biol Conserv. 1977;12(2):115–34.
47. Brydegaard M, Gebru A, Svanberg S. Super Resolution Laser Radar with Blinking Atmospheric Particles—Application to Interacting Flying Insects. Prog Electromagn Res. 2014;147:141–51.
48. Malmqvist E, Jansson S, Török S, Brydegaard M. Effective parameterization of laser radar observations of atmospheric fauna. IEEE J Sel Top Quantum Electron. 2015;22:1.
49. Rosenkranz BC, Lund J. Danmarks Højdemodel-én model med et utal af anvendelser. Geoforum Perspekt. 2015;14:26.
50. Mei L, Brydegaard M. Continuous-wave differential absorption lidar. Laser Photonics Rev. 2015;9(6):629–36.
51. Brydegaard M, Gebru A, Svanberg S. Super resolution laser radar with blinking atmospheric particles—application to interacting flying insects. Prog Electromagn Res. 2014;147:141–51.
52. Torok S. Kilohertz electro-optics for remote sensing of insect dispersal. These. 2013.
53. Malmqvist E. From Fauna to Flames : remote sensing with Scheimpflug-Lidar. [Lund]: Division of Combustion Physics, Department of Physics, Lund University; 2019.
54. Jansson S. Entomological lidar : target characterization and field applications. Lund: Division of Combustion Physics, Department of Physics, Lund University; 2020.

55. Page RE, Metcalf RA. A population investment sex ratio for the honey bee (*Apis mellifera* L.). Am Nat. 1984;124(5):680–702.
56. Allen MD. Drone production in honey-bee colonies (Apis mellifera L.). Nature. 1963;199(4895):789–90.
57. Loper GM, Wolf WW, Taylor OR. Honey bee drone flyways and congregation areas: radar observations. J Kansas Entomol Soc. 1992;1:223–30.
58. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72.
59. Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans audio Electroacoust. 1967;15(2):70–3.
60. Byrne, David N and Buchmann, Stephen L and Spangler HG. Relationship Between Wing Loading, Wingbeat Frequency and Body Mass in Homopterous Insects. J Exp Biol. 1988;135(1):9 LP – 23.
61. Feuerbacher E, Fewell JH, Roberts SP, Smith EF, Harrison JF. Effects of load type (pollen or nectar) and load mass on hovering metabolic rate and mechanical power output in the honey bee Apis mellifera. J Exp Biol. 2003;206(11):1855–65.
62. Altshuler DL, Dickson WB, Vance JT, Roberts SP, Dickinson MH. Short-amplitude high-frequency wing strokes determine the aerodynamics of honeybee flight. Proc Natl Acad Sci U S A. 2005;102(50):18213–8.
63. Timeanddate.com. [cited 2020 Apr 17]. https://www.timeanddate.com/sun/@2618263?month=7&year=2017. Accessed 17 Apr 2020.
64. Malmqvist E, Brydegaard M. Applications of KHZ-CW Lidar in Ecological Entomology. EPJ Web Conf. 2016;119:4–7.
65. Capaldi EA, Dyer FC. The role of orientation flights on homing performance in honeybees. J Exp Biol. 1999;202(12):1655–66.
66. Combes SA, Gagliardi SF, Switzer CM, Dillon ME. Kinematic flexibility allows bumblebees to increase energetic efficiency when carrying heavy loads. Sci Adv. 2020;6:6.
67. Tauc MJ, Fristrup KM, Repasky KS, Shaw JA. Field demonstration of a wing-beat modulation lidar for the 3D mapping of flying insects. OSA Contin. 2019;2(2):332.
68. Gao F, Lin H, Chen K, Chen X, He S. Light-sheet based two-dimensional Scheimpflug lidar system for profile measurements. Opt Express. 2018;26(21):27179.
69. Abou-Shaara H. The foraging behaviour of honey bees. Apis mellifera: A review Vet Med (Praha). 2014;1(59):1–10.
70. Hagler J, Mueller S, Teuber L, Machtley S, Deynze A. Foraging Range of Honey Bees, Apis mellifera, in Alfalfa Seed Production Fields. J Insect Sci. 2011;1(11):144.
71. Li M, Jansson S, Runemark A, Peterson J, Kirkeby CT, Jönsson AM, et al. Bark beetles as lidar targets and prospects of photonic surveillance. J Biophotonics. 2020.
72. Kirkeby C, Rydhmer K, Cook SM, Strand A, Torrance MT, Swain JL, et al. Advances in automatic identification of flying insects using optical sensors and machine learning. Sci Rep. 2021;11(1):1555.
73. Genoud AP, Gao Y, Williams GM, Thomas BP. A comparison of supervised machine learning algorithms for mosquito identification from backscattered optical signals. Ecol Inform. 2020;58:101090.

Publisher's Note

Paper II: Automating insect monitoring using unsupervised near-infrared sensors

scientific reports

Check for updates

OPEN

Automating insect monitoring using unsupervised near-infrared sensors

Klas Rydhmer[1,2]✉, Emily Bick[1,3], Laurence Still[1], Alfred Strand[1], Rubens Luciano[1], Salena Helmreich[1], Brittany D. Beck[1], Christoffer Grønne[1], Ludvig Malmros[1], Knud Poulsen[1], Frederik Elbæk[1], Mikkel Brydegaard[1,4,5,6], Jesper Lemmich[1] & Thomas Nikolajsen[1]

Insect monitoring is critical to improve our understanding and ability to preserve and restore biodiversity, sustainably produce crops, and reduce vectors of human and livestock disease. Conventional monitoring methods of trapping and identification are time consuming and thus expensive. Automation would significantly improve the state of the art. Here, we present a network of distributed wireless sensors that moves the field towards automation by recording backscattered near-infrared modulation signatures from insects. The instrument is a compact sensor based on dual-wavelength infrared light emitting diodes and is capable of unsupervised, autonomous long-term insect monitoring over weather and seasons. The sensor records the backscattered light at kHz pace from each insect transiting the measurement volume. Insect observations are automatically extracted and transmitted with environmental metadata over cellular connection to a cloud-based database. The recorded features include wing beat harmonics, melanisation and flight direction. To validate the sensor's capabilities, we tested the correlation between daily insect counts from an oil seed rape field measured with six yellow water traps and six sensors during a 4-week period. A comparison of the methods found a Spearman's rank correlation coefficient of 0.61 and a p-value = 0.0065, with the sensors recording approximately 19 times more insect observations and demonstrating a larger temporal dynamic than conventional yellow water trap monitoring.

Insecta is the most speciose class of terrestrial fauna[1] and the majority of the world's biodiversity is composed of this class[2]. In epidemiological and agricultural ecosystems, insects serve as both beneficial organisms[3–5] and economic pests[6,7]. Data on insects can support biodiversity conservation[8,9], human health protection[10] and increased food production[11].

Insects are monitored via established sampling methods including trapping, sweep netting, and portable aspiration[12–14]. These methods are imperfect resulting in biases towards size[15–17] and stage[18]. Additionally, conventional methods may be time-consuming, costly and prone to human error such as person-to-person variation in sampling execution[19–21]. New methods, like insect anesthetization sampling[22], are being implemented to minimize these biases. Regardless of sampling method, insect identification is time consuming and requires specialized training.

In order to reduce the cost of insect monitoring and identification, automation of insect trapping[23–27] and identification[27–31] has been developed. While these methods could greatly improve monitoring via traps, they are unsuitable for monitoring a general insect population since trap designs and baits are generally biased in regard to species[32,33].

Automation of insect monitoring without traps could reduce species bias of conventional methods and human error, thus greatly improving the state of the art. Insect identification has been automated as early as 1973 using wingbeat frequency[34–36], and today remote insect sensing includes acoustic detection[37], radar observations[38–40] and lidar[41–43]. Acoustic methods work best with a solid medium[26,44], though acoustic monitoring of free flying insects has been demonstrated[45–47]. While radar technologies have much larger monitoring range[16,40,48–50],

[1]FaunaPhotonics APS, Støberigade 14, 2450 Copenhagen, SV, Denmark. [2]Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark. [3]Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg C, Denmark. [4]Department of Physics, Lund Laser Centre, Lund University, Sölvegatan 14, 223 62 Lund, Sweden. [5]Department of Biology, Center for Animal Movement Research, Lund University, Sölvegatan 35, 223 62 Lund, Sweden. [6]Norsk Elektro Optikk AS, Østensjøveien 34, 0667 Oslo, Norway. ✉email: klry@faunaphotonics.com

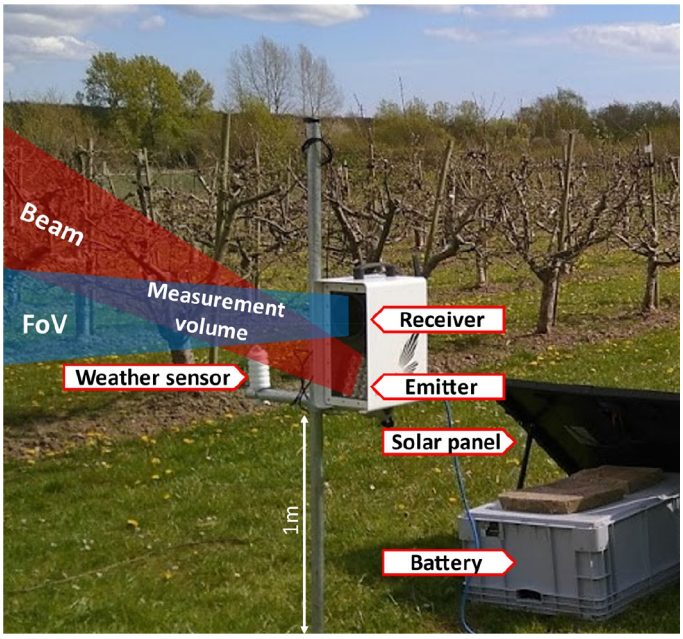


Figure 1. Situational photo of the sensor. As insects fly into the measurement volume, the backscattered light is recorded by the receiver. Insect observations are automatically extracted and transmitted along with environmental data, location, and situational photos, to the cloud via a GSM connection. Using a solar panel and battery, the sensor is capable of unsupervised, long-term monitoring in remote locations.

they are unsuitable for monitoring small insects, or insects around vegetation, such as a crop canopy. Optical methods were early used as to overcome many of these limitations[51–53]. Today, lidar can be used to record a large number of observations in a long transect[54–58] and distinguish between species groups by wingbeat frequency (WBF)[55,59]. However, lidar equipment requires a trained operator and requires constant supervision due to eye safety restrictions.

Here we present an autonomous near-infrared sensor for monitoring of flying insects in the field. The sensor aims to minimize human biases, be usable by non-technical personnel, and be capable of unsupervised long-term monitoring. Compared to existing entomological lidars, it has a smaller measurement volume but is eye safe and weatherproof.

Instrument design

The sensor is weatherproof, compact, and intended for field use by non-technicians. Like entomological lidar instrumentation, an air volume is illuminated, and light backscattered from insects entering the measurement volume is recorded by a high-speed photodetector. In addition, the instrument is equipped with a satellite navigation device, a camera for situational photos, and an environmental sensor monitoring temperature, humidity, and light intensity. An internal Global System for Mobile Communications (GSM) modem allows for communication and data transfer. The sensor can be powered by any 12 V power supply, including utility power, batteries, or solar power, and has a maximum power consumption of 30 W during monitoring. A photo of the sensor is shown in Fig. 1 and an internal block diagram is described in Fig. 2.

Emitter. The emitter module consists of a rectangular array of LEDs emitting two spectral bands at 808 nm and 980 nm with total output of 1.6 W and 1.7 W, respectively. The two wavelengths are modulated in a square wave at 118.8 kHz and 79.2 kHz respectively. The LEDs are mounted in a checkerboard pattern to achieve a homogeneous beam profile. The total area of the checkerboard, and thus the beam size at the source, is 82 cm^2. The light emitted from each diode is partially collimated by an asymmetrical lens and expands with 20° and 4° diverging angles (θ_E). The full width half maximum (FWHM) of the emitted light is 26 nm for the 808 nm band and 47 nm for the 970 band.

Receiver. The backscattered light from insects entering the overlap between the beam and the receiver's field of view (FoV) is collected by a near infrared coated aspheric lens (60 mm focal length, ø 76.2 mm aperture) onto a silicon quadrant photodiode (QPD) with a total area of 1 cm^2. The receiver is focused at 1 m and has a 4° divergence angle (θ_R). Quadrant detection of insects allow for basic range and size estimation[60,61] and can differentiate ascending and descending insects as well as migrating insects with tailwind or host- or scent-seeking insects with headwind.

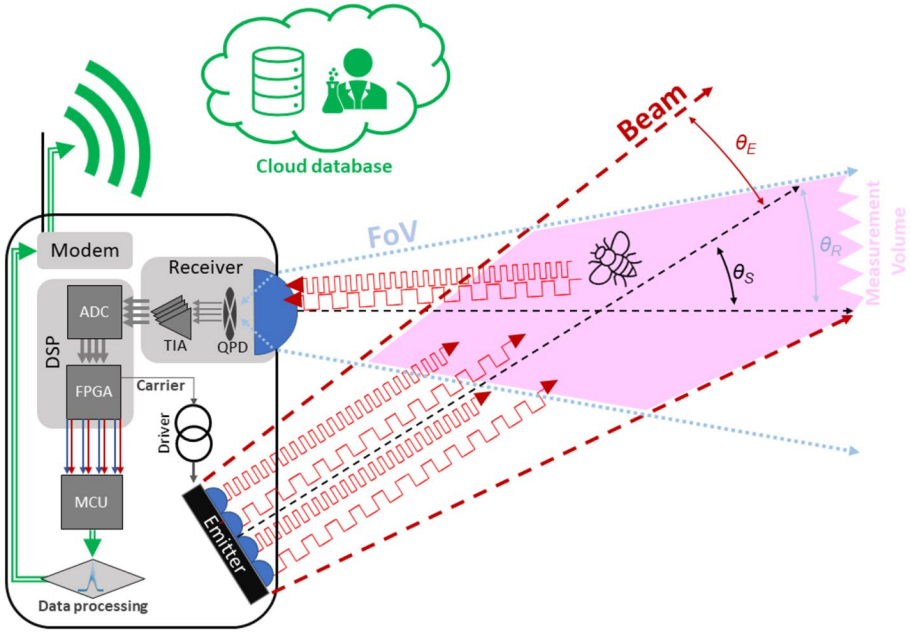


Figure 2. General measurement principle. Light is emitted and collimated from the LED board at 808 nm and 980 nm and modulated at different carrier frequencies. The backscattered light from an insect entering the measurement volume is collected by a lens and focused onto a QPD. The four QPD-quadrants are independently amplified by a TIA and sampled. The digital data streams are sent to the FPGA, where 8 digital lock-in amplifiers individually amplify each wavelength in the digital signal processing (DSP) unit. The resulting 8-channel data stream is analyzed by the MCU which extracts events from the data stream. The events can then be stored locally or sent via GSM modem to a cloud database. Created using Power Point 365.

Signal processing. Each quadrant of the QPD is amplified by a dedicated trans-impedance amplifier (TIA) with a bandwidth of 10 Hz–1 MHz and a gain of 0.75 V/μA around 100 kHz. The amplified signals are sampled by four analogue–digital converters (ADC) with 14-bit output at a rate of 6 MHz. The digital data-streams are sent into a field-programmable gate array (FPGA) where eight digital lock-in amplifiers are implemented in VHDL (Very High-Speed Integrated Circuit Hardware Description Language). This allows the two spectral bands to be recorded independently on each quadrant, resulting in an 8-channel data stream. The data is then filtered by a low-pass filter with a cut-off at 5 kHz and digitally sampled to a 20 kHz, 16-bit data stream before it is sent to a microcontroller unit (MCU) for event extraction and further processing (Fig. 3). Since insects generally have wing beat frequencies below 1 kHz, a 5 kHz cutoff allows us to resolve a minimum of five harmonics in the frequency spectra. The increase in bit depth is possible due to the oversampling of the unfiltered signal.

Measurement volume. The measurement volume is defined by the overlap between the beam and the FoV. Its size and shape can be adjusted by changing the angle (θ_S) between the emitter and receiver.

The beam, FoV and the measurement volume have been mapped by a custom-built 3-axis robot covering a volume of 2 m × 1.5 m × 1.5 m. The robot is equipped with a photodetector, an illumination source, and a sphere dropping mechanism. The photo-detector and illumination source are used to map the emitted beam and FoV respectively while sphere dropping mechanism allow us to verify the signal intensity from a standard object at any point in the volume. Using these methods, the signal response from an arbitrary target can be estimated. The volumes were measured at 20 planes along the Z axis, from 30 to 1655 mm, each plane consisting of 56 × 56 measurement points in a 12 mm grid. The calculated signals were then compared to actual measurement values by dropping black and white spheres. The white spheres were assumed to be 100% reflective and the black spheres had a 5% reflectivity.

The measurement volume properties for targets with various optical cross sections (OCS) at different angles are shown in Table 1. The size of the measurement volume is dependent on the minimum acceptable sensitivity, which is related to the noise in the instrument. In the following results, the edge of the volume is defined as the limit where the signal to noise ratio (SNR) is larger than 10 for typical noise levels in a field installation. The signal to noise ratio is defined as the maximum value of the recorded signal divided by the peak-to-peak noise. The volumes for a 10 mm^2 target are shown in Fig. 4.

Data processing. *Automated event extraction.* The sensor records intervals of 10 min (4 quadrants, 2 spectral bands, 16 bit and 20 kHz sample rate after demux of carrier frequency) and automatically extracts insect observations from each recording. The event extraction is inspired by earlier work but modified to reduce computational load[42,43,59,62]. The event extraction algorithm was developed during prior experiments in various

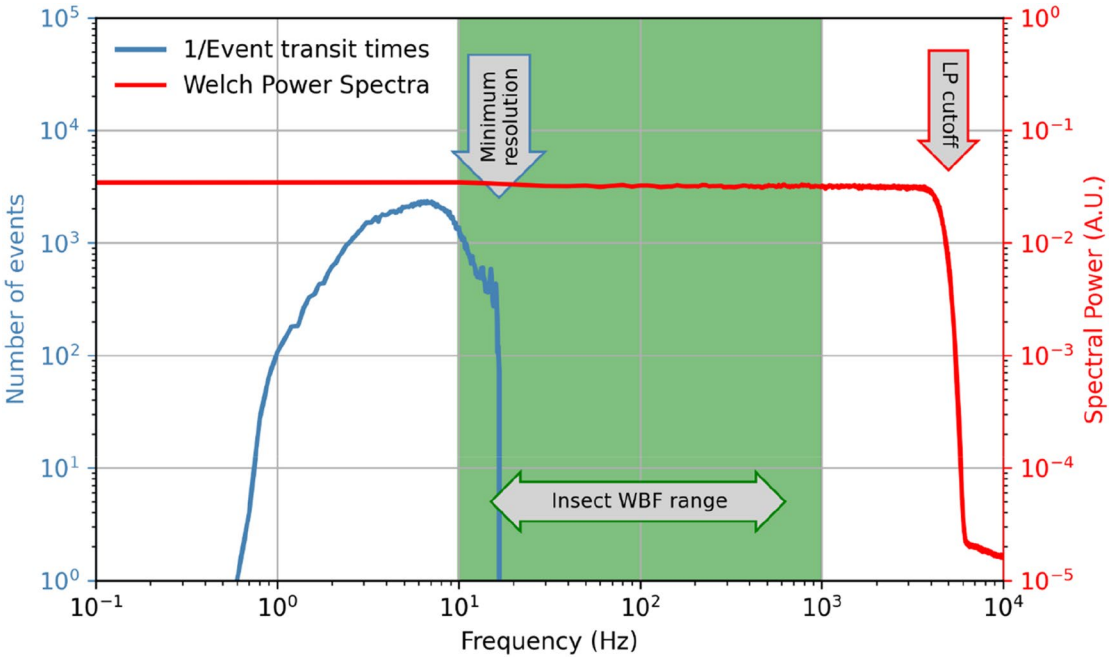


Figure 3. Frequency diagram. The wide beam yields long insect transit times, and the corresponding frequency resolution is high enough to accurately capture most species. The frequency response curve (red) is flat in the wingbeat frequency region and the effect of the LP filter at 5 kHz is indicated. The 5 kHz bandwidth allows a minimum of 4 harmonic overtones to be recorded even for mosquitoes with very high wingbeat frequencies.

θ_S (deg)	SNR at 25 cm for 10 mm^2 target (dB)	Far limit (10 mm^2 target) (cm)	Measurement volume for 1 mm^2 target (L)	Measurement volume for 10 mm^2 target (L)	Measurement volume for 100 mm^2 target (L)
5	28.6	>1650	8	52	100
12.5	31.5	130	7	27	87
20	32.3	95	5	16	70

Table 1. Measurement volume parameters at different angles for different target OCS. The target OCS values correspond roughly to a small midge, a small beetle, and a honeybee.

conditions. In simple terms, it aims to quantify the noise level and subsequently multiply it with a signal-to-noise factor to yield a threshold. All events that exceed this threshold are then extracted.

In the chosen implementation, the signal in each channel was downsampled to 2 kHz and a rolling median boxcar filter with a width of 2 s and 50% overlap was used to estimate the quasi-static baselines (the baselines can change with environmental conditions, static objects in the beam etc.). A 2 s window width makes the median estimation insensitive to insect observations, which has an average transit time of ca 100 ms. The standard deviation of the baseline was measured with an identical filter, applied to all datapoints below the median. The selection of values below the median reduces the influence of rare events, such as insects, on the noise level estimation.

The interpolated median signals were removed from the full resolution data and we employed a Boolean condition for insect detection when the time series exceed ten times the estimated standard deviation. A high threshold factor rejected weak observations which could yield unreliable results in the downstream feature extraction. The Boolean time series were eroded by 500 µs and dilated by 30 ms. The erosion rejects short spikes, outliers and insect signals to short to be interpreted and the dilation includes insect observation flanks. The logical OR function was applied across all QPD-quadrants and spectral channels. Extracted observations are transmitted to a cloud database along with metadata such as baseline and noise level, via GSM connection or stored locally until a connection is available. An example of the event extraction process is shown in Fig. 5, and the insect event is shown in greater detail in Fig. 6.

Each insect observation, along with its associated timestamp and device identifier, is automatically uploaded to the cloud via one-way AMQP (Advanced Message Queuing Protocol), with unique connections for each device. Virtual computing is then used to further process, analyze, and securely store data for further use and aggregation.

Feature extraction/data interpretation. The QPD segments collect backscattered light from different sections of the measurement volume. For a single object passing through the measurement volume, the signal strength within each QPD-quadrant is related to the object's OCS as well as its position. As the OCS varies with each wingbeat, the wingbeat frequency can be resolved. Many methods have been used to extract the wingbeat

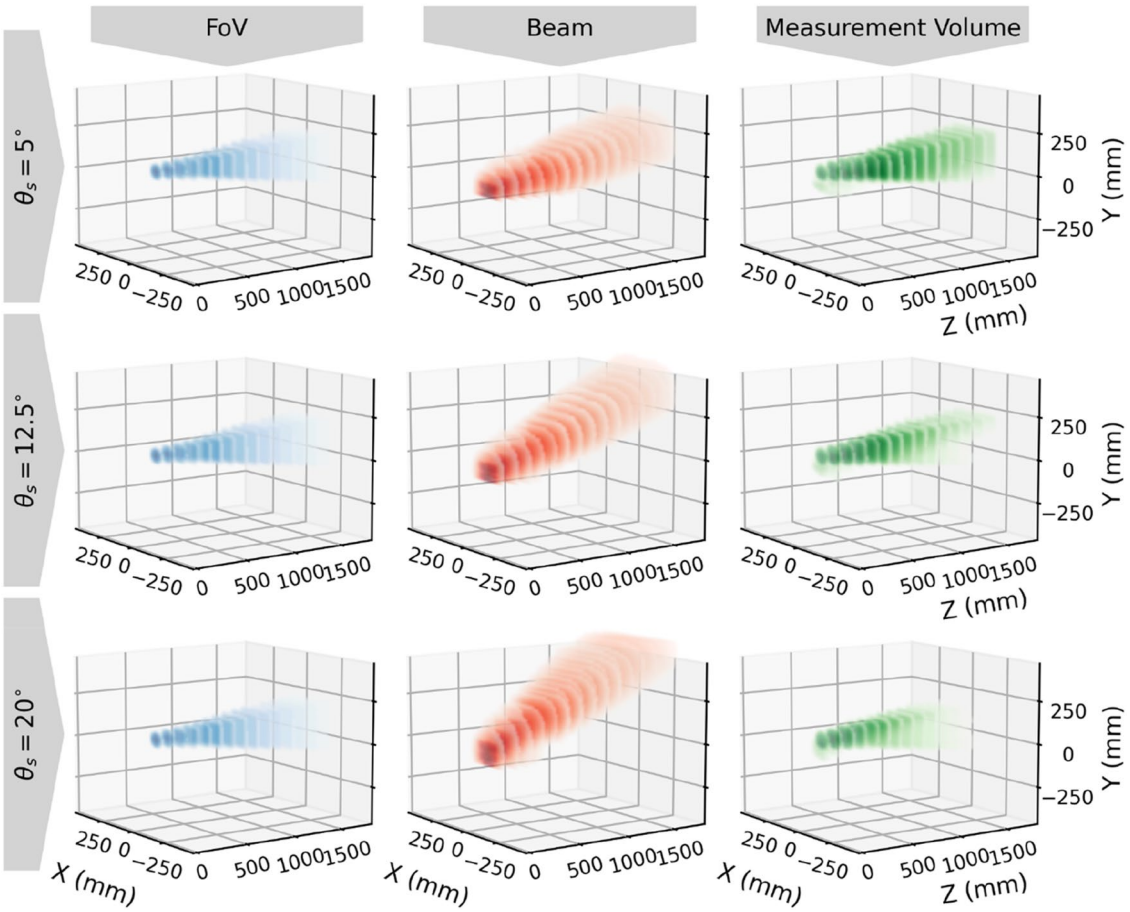


Figure 4. Measured FoV, beam, and measurement volume for the three angles. Each volume is mapped at 20 planes along the Z axis and each plane consists of 56 × 56 measurement points with 12 mm spacing. For the FoV and beam, all measurement points below 2% of the maximum value are excluded. For the measurement volume all points with a SNR < 10 for a 10 mm^2 target are excluded. A low angle yields a longer and larger, but less sensitive, measurement volume. The FoV is identical in all configurations.

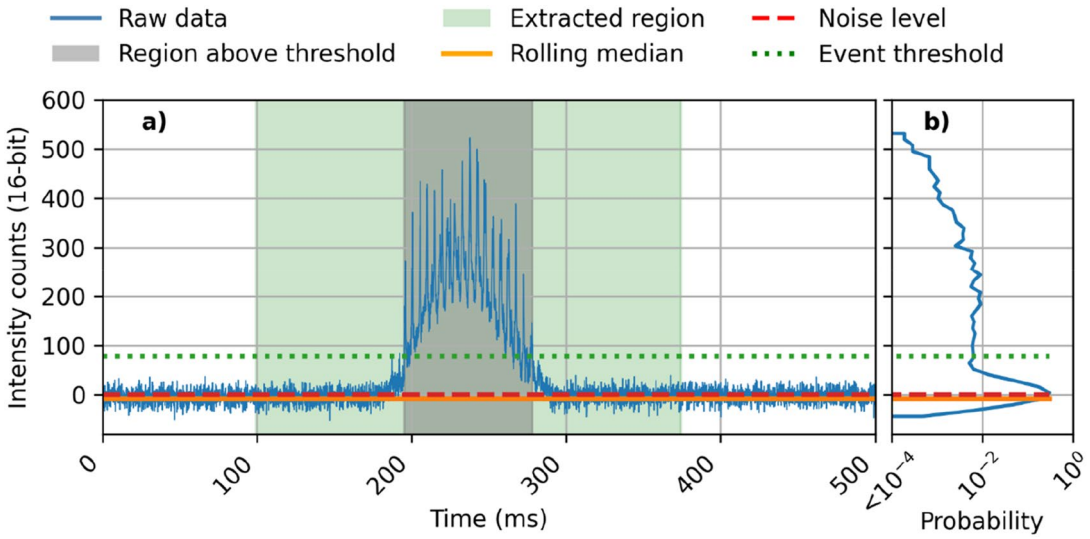


Figure 5. An example of the event extraction process in a single channel for visibility. (**a**) The data, in the 810 nm band of a single QPD segment after the rolling median has been removed. The part of the signal above the event threshold is marked in grey, and the final insect event after erosion and dilation of the binary map is marked in green. (**b**) Intensity distribution of the data.

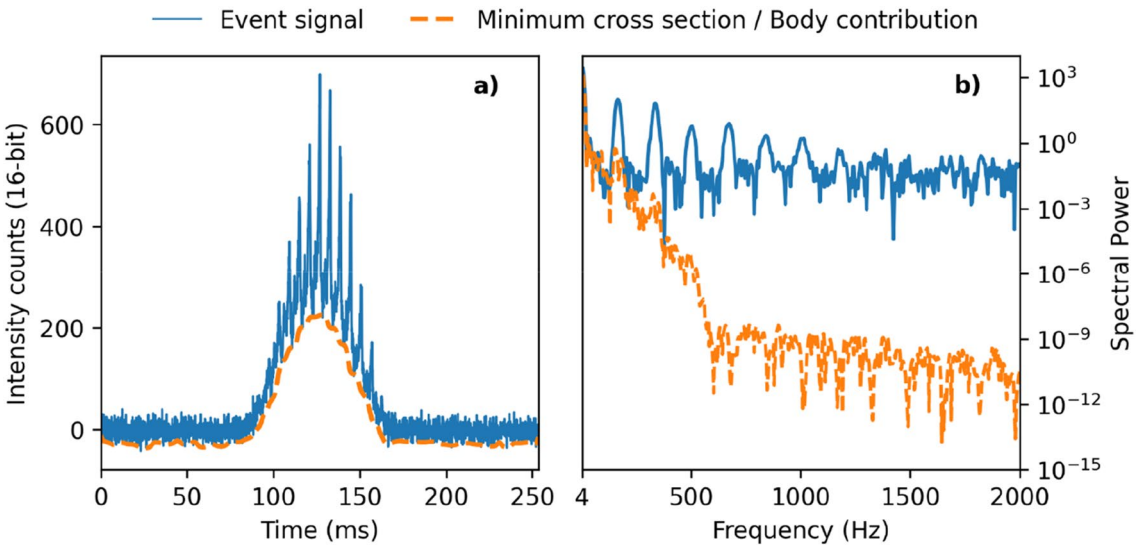


Figure 6. Insect event example. (**a**) The 810 nm signal for a single insect event in of one of the QPD segments. The insect wingbeats appear as undulating spikes. The minimum envelope of the signal is interpreted as the insect body contribution to the signal. (**b**) The Welch spectral density of the event. The fundamental wingbeat frequency and harmonics are seen in the event signal. This event has a fundamental wingbeat frequency of 160 Hz and an average body-to-wing ratio of 0.4.

frequency from insect observations[62–64] and most are based on identifying the fundamental frequency in the frequency domain, as shown in Fig. 6b.

In addition to the wingbeat frequency, the body and wing contribution can be measured from each time signal which allows calculation of additional features such as body-to-wing ratio. Additional features can be calculated by comparing the relative intensity of the body and wing signals in the two spectral bands. These bands differentially index melanin absorption[65–67] and may yield some sensitivity to wing interference patterns[66,68,69], although not enough to uniquely determine wing membrane thickness. Together these features can be used to quantify the morphology of different insect groups and allow remote classification of insects according to order, family, genus or species[32,64,69,70].

Field validation

Methodology. The sensor was field-tested against a conventional insect monitoring method, yellow water traps (22 cm diameter)[33,71], in an organic oilseed rape (*Brassica napus* L.) field in the vicinity of Sorø, Denmark (55° 29′ 04.3″ N 11° 29′ 34.6″ E). During a four-week period (04/22/20–05/22/20), insects were monitored with six sensors and six yellow water traps. The water traps were filled with water and soap, immediately drowning any insects landing in the trap. Sensors and traps were placed in a grid pattern, consisting of four linear transects 30 m from and perpendicular to the field's southern-most edge. This is illustrated in Fig. 7. Each transect consisted of three monitoring points (either sensors or traps) with 45 m spacing, and a separation of 22.5 m between transects. The first and third transect consisted of sensors and the second and fourth were yellow water traps. During the field study presented in this work, θ_S was set to 20° in order to maximize the signal strength of small targets at close range.

Fundamentally the two methods observe different insect behaviors. While the sensor looks at insects flying above the crop canopy, the yellow water traps look at insects that occur within it. Further confounding the comparison, yellow is attractive to some insects[33]. Therefore, some proportion of insects will be attracted to the yellow water traps, resulting in overrepresentation of some species[72,73]. However, water traps constitute the standard practice for pest monitoring in oilseed rape for many species.

Data analysis. The water traps were emptied daily, except for Sundays and 7 additional days (3 sample days in late April and 4 days in mid May) where we were unable to empty the traps. Sensor data was recorded continuously. All insects in the traps were collected, but to allow for a more direct comparison of methods, non-flying insects and thrips found in water traps were excluded from further analysis.

The sensor data was aggregated according to the collection time of the water traps. Insects trapped during Sundays were added to the following days count and the number of collected insects was normalized by the number of trapping days. One day, April 30th, was excluded due to instrument malfunction. The average number of recorded insect observations per sensor per day and per hour was calculated. The calculated numbers were normalized by sensor uptime, which was on average 90% throughout the measurement period. Observations during heavy rainfall and without any distinguishable wingbeat frequency, ca 1% of the observations, were automatically removed from the data using a classification algorithm.

Figure 7. Layout of sensors and traps on the field. Sensors and traps were placed in a grid pattern ca 30 m from the field edges. The four north–south transects are separated by ca 22 m and consists of either sensors or water traps, spaced by 45 m. Image data from Google Earth 2021, Aerodata International Survey Mapdata 2021.

Results

The insect activity recorded by the sensors and traps respectively are shown in Fig. 8. Insect counts from sensors and traps cannot be directly equated due to differences in measurement subject (insect flights vs insect landings) and non-homogeneous insect distribution; however, they serve to visualize similarities in gross changes in insect activity over the sample period. The results demonstrate a significant correlation between the sensor and trap results, specifically with a Spearman's rank correlation coefficient of 0.61 and a p-value = 0.0065[74]. Over the course of the season, an average of 1122 ± 242 (SE) insect observations per day were collected per sensor (excluding downtime), compared to an average of 63 ± 6 (SE) insects caught per water trap per day over the same period.

Discussion

Here we present a sensor for automated unsupervised field monitoring of insect flight activity. The sensor illuminates an air volume and records the backscattered light from insects that fly through the measurement volume. Discrete insect observations are automatically extracted from the continuous raw data flow and transmitted over a cellular connection to a database in the cloud. Field validation showed the number of recorded insect observations correlates with the number of individual insects trapped by a conventional insect monitoring method. Furthermore, the sensor recorded an order of magnitude more insects than the conventional method over the same period.

The automation of insect monitoring has the potential to reduce monitoring bias, cost, and human labor, potentially resulting in an increased ability to collect large quantities of biodiversity, public health, and economically relevant insect data. Additionally, the observations from the sensors were available in real time, whereas emptying and counting insects from traps required a significant amount of labor. While this work was limited to comparing total insect counts from the traps, it is possible for a skilled expert to identify these insects to the sub-species level. This is an area were the traps currently have a strong advantage over this sensor and similar instrumentation. Developing and evaluating species specific insect classification algorithms is therefore a major focus. Significant work is still needed prior to field implementation to test possible use cases and limitations of this system.

One of the most striking differences in monitoring methods is the day-to-day variability in the number of data points collected (Fig. 7). While the yellow traps catch a similar number of insects each day, the difference between low and high flight activity days were more visible in the sensor. Early analysis of the trap and sensor

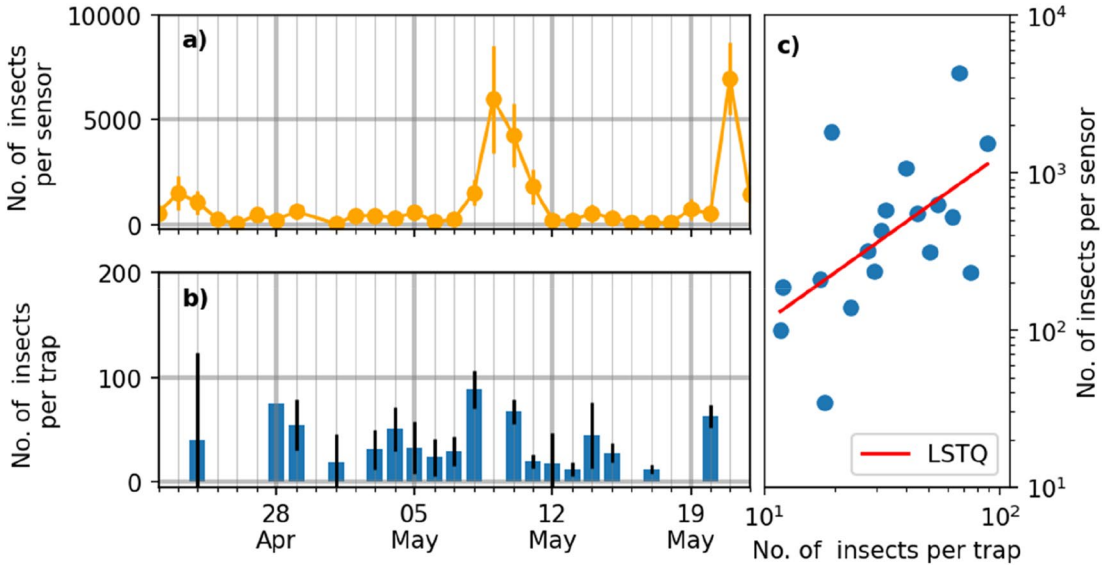


Figure 8. Sensor-trap comparison. (**a**) Average insect counts across sensors per day. Errorbars indicate the standard deviation between the sensors. (**b**) Average insect count across yellow traps per day. Errorbars indicate the standard deviation between the traps. (**c**) Sensor vs trap counts during days where both sensor and trap data were available. The red line is the linear least square fit (LSTQ) with a Spearman correlation coefficient of 0.61.

data indicates that the peak recorded during May 7–11 is due to a pollen beetle (*Brassicogethes aeneus*) activity spike. This will be the subject of further studies.

Another marked difference between the sensor and the water traps is the number of data points collected over the same collection period. Each sensor observed ~ 19× more insect observations than insects collected in the water trap. While in general the correlation between the two values is considered more relevant than the absolute number, one advantage of a much higher observation is the ability to get statistically sound data aggregated with very high temporal resolution. In this work, the data was aggregated to match the collection times of the traps but it could easily be aggregated down to hourly activity. The higher temporal resolution and continuous monitoring during unsociable hours allows for the comparatively easy and low-labor collection of data on insect circadian rhythms, as well as direct weather interactions.

We hypothesize that the sensors observe different insect behaviors compared to conventional monitoring methods since only airborne (flying or jumping) insects are recorded. Therefore, we did not expect a perfect correlation between the sensors and the conventional methods. Sweep netting is likely the most similar monitoring method since it also catches insects in flight above the crop. However, sweep netting, which also collects insects on plants, occurs at a point measurement in time and is typically performed along a transect, rather than at a fixed point in the field[19]. Also, each trapping method is biased towards different insects, influencing catch[15,17].

Trapping methods, such as the water traps used in this study, monitor insects landing, walking, or jumping to a specific point and do not record insects in flight. Also, each trapping method is biased towards different insects, with the trap color influencing the trap catch[33]. It would therefore be beneficial to include multiple trap types in the ground truthing in future work. Additionally, both the sensors and the conventional ground truthing methods assume that the recorded insect activity in one specific point in the field is representative of the insect activity in the near surroundings.

Although we do not fully understand in what manner, the sensor is also most likely biased towards reporting certain species groups. Most primarily, its only capable of recording airborne insects and unsuitable for monitoring during rain. Insect vision is focused towards the visual or ultraviolet spectrum and not capable of resolving infrared light and we believe the emitted beam has very little influence on insect behavior[75]. However, in a homogeneous landscape such as an agricultural field, any foreign object placed above the canopy could serve as an attractant to insects. Finally, the size of the measurement volume varies with the OCS of the insects and larger insects will be over-represented. To provide a complete picture of the insect population, this should be considered. Along with species specific observations, this is an area where we expect significant progress.

Automated insect monitoring has the potential to facilitate pest prevention, public health studies and biodiversity monitoring. Compared to alternative methods, such as automated traps, the sensor described in this paper comes at a higher cost per unit and with higher power requirements. Compared to previously described entomological lidars, we record fewer observations, but with a longer transit time and higher sampling frequency. We believe the advantage of entomological lidars, such as the sensor described in this paper, is the ability to potentially monitor and discriminate between multiple species using a single instrument.

In further work we will explore the possibilities of unsupervised long-term monitoring of insect activity and species recognition.

Conclusions

In this work, we have introduced an unsupervised automated sensor for insect monitoring. The measurement principle is similar to entomological lidar setups but is optimized for near-field measurements. This simplifies the installation process and increases the robustness of the sensor, allowing it to be operable by non-technical experts and enables long-term unsupervised monitoring.

The sensor automatically extracts insect events from the raw data and transmits these via a built-in modem for further processing. From the recorded observations, features such as the wingbeat frequency, body-wing ratio, and melanisation factor are computed and used to predict the insect classification down to species. During a 4-week deployment in an oilseed rape field, the detected flight activity was shown to be correlated with a conventional monitoring method.

The capabilities, standardization, and scalability of this sensor-based method has the potential to improve the state of the art in insect monitoring. To date, 119 similar units have been deployed in field and in 2021 the cloud database encompassed > 18 million insect observations. The sensor can be used to explore areas such as biodiversity assessment, insecticide resistance, and long-term monitoring of remote areas, facilitating research studies currently difficult or impossible to conduct with conventional methods.

Received: 17 August 2021; Accepted: 28 January 2022
Published online: 16 February 2022

References

1. Stork, N. E. How many species of insects and other terrestrial arthropods are there on earth? (2017). https://doi.org/10.1146/annurev-ento-020117.
2. Scudder, G. *Insect Biodiversity: Science and Society—Google Books* (Wiley-Blackwell, 2009).
3. Lami, F., Boscutti, F., Masin, R., Sigura, M. & Marini, L. Seed predation intensity and stability in agro-ecosystems: Role of predator diversity and soil disturbance. *Agric. Ecosyst. Environ.* **288**, 106720 (2020).
4. Gallai, N., Salles, J. M., Settele, J. & Vaissière, B. E. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.* **68**, 810–821 (2009).
5. Consoli, F. L., Parra, J. R. P. & Zucchi, R. A. *Egg Parasitoids in Agroecosystems with Emphasis on Trichogramma* (Springer Science, 2010).
6. Sánchez-Guillén, R. A., Córdoba-Aguilar, A., Hansson, B., Ott, J. & Wellenreuther, M. Evolutionary consequences of climate-induced range shifts in insects. *Biol. Rev.* **91**, 1050–1064 (2016).
7. Zalucki, M. P. *et al.* Estimating the economic cost of one of the world's major insect pests, *Plutella xylostella* (Lepidoptera: Plutellidae): Just how long is a piece of string?. *J. Econ. Entomol.* **105**, 1115–1129 (2012).
8. Dornelas, M. & Daskalova, G. N. Nuanced changes in insect abundance. *Science (80-).* **368**, 368–369 (2020).
9. Didham, R. K. *et al.* Interpreting insect declines: Seven challenges and a way forward. *Insect Conserv. Divers.* **13**, 103–114 (2020).
10. Greenwood, B. M., Bojang, K. & Whitty, C. J. M. Malaria. *Lancet* **365**, 98 (2005).
11. Dangles, O. & Casas, J. Ecosystem services provided by insects for achieving sustainable development goals. *Ecosyst. Serv.* **35**, 109–115 (2019).
12. Burkholder, W. E. & Ma, M. Pheromones for monitoring and control of stored-product insects. *Annu. Rev. Entomol.* **30**, 257–272 (1985).
13. Morris, R. F. Sampling insect populations. *Annu. Rev. Entomol.* **5**, 243–264 (1960).
14. Strickland, A. H. Sampling crop pests and their hosts. *Annu. Rev. Entomol.* **6**, 201–220 (1961).
15. Bannerman, J. A., Costamagna, A. C., McCornack, B. P. & Ragsdale, D. W. Comparison of relative bias, precision, and efficiency of sampling methods for natural enemies of soybean aphid (Hemiptera: Aphididae). *J. Econ. Entomol.* **108**, 1381–1397 (2015).
16. Osborne, J. L. *et al.* Harmonic radar: A new technique for investigating bumblebee and honey bee foraging flight. *VII Int. Symp. Pollinat.* **437**, 159–164 (1996).
17. Zink, A. G. & Rosenheim, J. A. State-dependent sampling bias in insects: Implications for monitoring western tarnished plant bugs. *Entomol. Exp. Appl.* **113**, 117–123 (2004).
18. Rancourt, B., Vincent, C. & De Oliveira, A. D. Circadian activity of *Lygus lineolaris* (Hemiptera: Miridae) and effectiveness of sampling techniques in strawberry fields. *J. Econ. Entomol* **93**, 1160–1166 (2000).
19. Binns, M. R. & Nyrop, J. P. Sampling insect populations for the purpose of IPM decision making. *Annu. Rev. Entomol.* **37**, 427–453. https://doi.org/10.1146/annurev.ento.37.1.427 (1992).
20. Portman, Z. M., Bruninga-Socolar, B. & Cariveau, D. P. The state of bee monitoring in the United States: A call to refocus away from bowl traps and towards more effective methods. *Ann. Entomol. Soc. Am.* **113**, 337–342 (2020).
21. Montgomery, G. A., Belitz, M. W., Guralnick, R. P. & Tingley, M. W. Standards and best practices for monitoring and benchmarking insects. *Front. Ecol. Evolut.* **8**, 579193 (2021).
22. Bick, E., Dryden, D. M., Nguyen, H. D. & Kim, H. A novel CO_2-based insect sampling device and associated field method evaluated in a strawberry agroecosystem. *J. Econ. Entomol.* **113**, 1037–1042 (2020).
23. Wen, C. & Guyer, D. Image-based orchard insect automated identification and classification method. *Comput. Electron. Agric.* **89**, 110–115 (2012).
24. Chen, Y., Why, A., Batista, G., Mafra-Neto, A. & Keogh, E. Flying insect classification with inexpensive sensors. *J. Insect Behav.* **27**, 657–677 (2014).
25. Potamitis, I. & Rigakis, I. Novel noise-robust optoacoustic sensors to identify insects through wingbeats. *IEEE Sens. J.* **15**, 4621–4631 (2015).
26. Eliopoulos, P. A., Potamitis, I., Kontodimas, D. C. & Givropoulou, E. G. Detection of adult beetles inside the stored wheat mass based on their acoustic emissions. *J. Econ. Entomol.* **108**, 2808–2814 (2015).
27. Ärje, J. *et al.* Automatic image-based identification and biomass estimation of invertebrates. *Methods Ecol. Evol.* **11**, 922–931 (2020).
28. Hobbs, S. E. & Hodges, G. An optical method for automatic classification and recording of a suction trap catch. *Bull. Entomol. Res.* **83**, 47–51 (1993).
29. O'Neill, M. A., Gauld, I. D., Gaston, K. J. & Weeks, P. Daisy: An automated invertebrate identification system using holistic vision techniques. in *Proceedings of the Inaugural Meeting BioNET-INTERNATIONAL Group for Computer-Aided Taxonomy (BIGCAT)* 13–22 (1997).
30. Chesmore, E. D. Methodologies for automating the identification of species. in *First BioNet-International Work. Gr. Autom. Taxon.* 3–12 (2000).
31. Martineau, M. *et al.* A survey on image-based insect classification. *Pattern Recognit.* **65**, 273–284 (2017).

32. Silva, D. F., De Souza, V. M. A., Batista Geapa, K. E. & Ellis, D. P. W. Applying machine learning and audio analysis techniques to insect recognition in intelligent traps. in *Proceedings—2013 12th International Conference on Machine Learning and Applications, ICMLA 2013*. (2013).
33. Capinera, J. L. & Walmsley, M. R. Visual responses of some sugarbeet insects to sticky traps and water pan traps of various colors. *J. Econ. Entomol.*, **71**(6), 926–927 (1978).
34. Moore, A., Miller, J. R., Tabashnik, B. E. & Gage, S. H. Automated identification of flying insects by analysis of wingbeat frequencies. *J. Econ. Entomol.* **79**, 1703–1706 (1986).
35. Riley, J. R. Angular and temporal variations in the radar cross-sections of insects. *Proc. Inst. Electr. Eng. (IET)* **120**, 1229–1232 (1973).
36. Reed, S. C., Williams, C. M. & Chadwick, L. E. Frequency of wing-beat as a character for separating species races and geographic varieties of Drosophila. *Genetics* **27**, 349 (1942).
37. Mankin, R. W., Hagstrum, D. W., Smith, M. T., Roda, A. L. & Kairo, M. T. K. Perspective and promise: a century of insect acoustic detection and monitoring. *Am. Entomol.* **57**(1), 30–44 (2011).
38. Drake, V. A. & Reynolds, D. R. *Radar Entomology: Observing Insect Flight and Migration* (Cabi, 2012).
39. Long, T. *et al.* Entomological radar overview: System and signal processing. *IEEE Aerosp. Electron. Syst. Mag.* **35**, 20–32 (2020).
40. Drake, V. A., Hatty, S., Symons, C. & Wang, H. Insect monitoring radar: Maximizing performance and utility. *Remote Sens.* **12**, 596 (2020).
41. Brydegaard, M. & Jansson, S. Advances in entomological laser radar. *IET Int. Radar Conf.* https://doi.org/10.1049/joe.2019.0598 (2018).
42. Jansson, S. *Entomological Lidar: Target Characterization and Field Applications* (Department of Physics, Lund University, 2020).
43. Malmqvist, E. *From Fauna to Flames: Remote Sensing with Scheimpflug-Lidar* (Department of Physics, Lund University, 2019).
44. Mankin, R. W., Hagstrum, D. W., Smith, M. T., Roda, A. L. & Kairo, M. T. K. Perspective and promise: A century of insect acoustic detection and monitoring. *Am. Entomol.* **57**, 30–44 (2011).
45. Miller-Struttmann, N. E., Heise, D., Schul, J., Geib, J. C. & Galen, C. Flight of the bumble bee: Buzzes predict pollination services. *PLoS ONE* **12**, 1–14 (2017).
46. Li, Y. *et al.* Mosquito detection with low-cost smartphones: Data acquisition for malaria research. arXiv:1711.06346 [stat.ML] (2017).
47. Mukundarajan, H., Hol, F. J. H., Castillo, E. A., Newby, C. & Prakash, M. Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. *Elife* **6**, 1–26 (2017).
48. Osborne, J. L. *et al.* A landscape-scale study of bumble bee foraging range and constancy, using harmonic radar. *J. Appl. Ecol.* **36**, 519–533 (1999).
49. Smith, A. D., Riley, J. R. & Gregory, R. D. A method for routine monitoring of the aerial migration of insects by using a vertical-looking radar. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **340**, 393–404 (1993).
50. Chapman, J. W., Smith, A. D., Woiwod, I. P., Reynolds, D. R. & Riley, J. R. Development of vertical-looking radar technology for monitoring insect migration. *Comput. Electron. Agric.* **35**(2–3), 95–110 (2002).
51. Schaefer, G. W. & Bent, G. A. An infra-red remote sensing system for the active detection and automatic determination of insect flight trajectories (IRADIT). *Bull. Entomol. Res.* **74**, 261–278 (1984).
52. Farmery, M. J. Optical studies of insect flight at low altitude. (Doctoral dissertation, University of York, 1981).
53. Farmery, M. J. The effect of air temperature on wingbeat frequency of naturally flying armyworm moth (*Spodoptera exempta*). *Entomol. Exp. Appl.* **32**, 193–194 (1982).
54. Malmqvist, E. & Brydegaard, M. Applications of KHZ-CW lidar in ecological entomology. *EPJ Web Conf.* **119**, 25016. https://doi.org/10.1051/epjconf/2016I11925016 (2016).
55. Brydegaard, M. *et al.* Lidar reveals activity anomaly of malaria vectors during pan-African eclipse. *Sci. Adv.* **6**, eaay5487 (2020).
56. Malmqvist, E. *et al.* The bat–bird–bug battle: Daily flight activity of insects and their predators over a rice field revealed by high-resolution Scheimpflug Lidar. *Roy. Soc. Open Sci.* **5**(4), 172303 (2018).
57. Fristrup, K. M., Shaw, J. A. & Tauc, M. J. Development of a wing-beat-modulation scanning lidar system for insect studies. *Lidar Remote Sens. Environ. Monit.* **2017**, 15. https://doi.org/10.1117/12.2274656 (2017).
58. Hoffman, D. S., Nehrir, A. R., Repasky, K. S., Shaw, J. A. & Carlsten, J. L. Range-resolved optical detection of honeybees by use of wing-beat modulation of scattered light for locating land mines. *Appl. Opt.* **46**, 3007–3012 (2007).
59. Jansson, S., Malmqvist, E. & Mlacha, Y. Real-time dispersal of malaria vectors in rural Africa monitored with lidar. *Plos one.* **16**(3), e0247803 (2021).
60. Jansson, S. & Brydegaard, M. Passive kHz lidar for the quantification of insect activity and dispersal. *Anim. Biotelemet.* **6**, 6 (2018).
61. Jansson, S. P. & Sørensen, M. B. An optical remote sensing system for detection of aerial and aquatic fauna. U.S. Patent Application No. 16/346,322 (2019).
62. Malmqvist, E., Jansson, S., Török, S. & Brydegaard, M. Effective parameterization of laser radar observations of atmospheric fauna. *IEEE J. Sel. Top. Quant. Electron.* **22**, 1 (2015).
63. Drake, V. A., Wang, H. K. & Harman, I. T. Insect Monitoring Radar: Remote and network operation. *Comput. Electron. Agric.* **35**, 77–94 (2002).
64. Kirkeby, C. *et al.* Advances in automatic identification of flying insects using optical sensors and machine learning. *Sci. Rep.* **11**, 1555 (2021).
65. Jacques, S. L. Erratum: Optical properties of biological tissues: A review (Physics in Medicine and Biology (2013) 58). *Phys. Med. Biol.* **58**, 5007–5008 (2013).
66. Li, M. *et al.* Bark beetles as lidar targets and prospects of photonic surveillance. *J. Biophoton.* https://doi.org/10.1002/jbio.202000420 (2020).
67. Brydegaard, M. Advantages of shortwave infrared LIDAR entomology. in *Laser Applications to Chemical, Security and Environmental Analysis* LW2D-6 (Optical Society of America, 2014).
68. Brydegaard, M., Jansson, S., Schulz, M. & Runemark, A. Can the narrow red bands of dragonflies be used to perceive wing interference patterns? *Ecol. Evol.* 8(11), 5369–5384 (2018).
69. Gebru, A. *et al.* Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight. *J. Biophotonics* **11**(8), e201800014 (2018).
70. Potamitis, I. Classifying insects on the fly. *Ecol. Inform.* **21**, 40–49 (2014).
71. Heathcote, G. D. The comparison of yellow cylindrical, flat and water traps, and of Johnson suction traps, for sampling aphids. *Ann. Appl. Biol.* **45**, 133–139 (1957).
72. Vaishampayan, S. M., Kogan, M., Waldbauer, G. P. & Woolley, J. Spectral specific responses in the visual behavior of the greenhouse whitefly, *Trialeurodes vaporariorum* (Homoptera: Aleyrodidae). *Entomol. Exp. Appl.* **18**, 344–356 (1975).
73. Mound, L. A. Studies on the olfaction and colour sensitivity of *Bemisia tabaci* (Genn.) (Homoptera, Aleyrodidae). *Entomol. Exp. Appl.* **5**, 99–104 (1962).
74. Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
75. Van Der Kooi, C. J., Stavenga, D. G., Arikawa, K., Belušič, G. & Kelber, A. Evolution of insect color vision: From spectral sensitivity to visual ecology. *Annu. Rev. Entomol.* **66**, 435–461 (2021).

Acknowledgements
The authors want to thank Jakob Dyhr for kindly making his organic oilseed rape field in Sorø, Denmark, available for the field experiments. Thanks to Lene Sigsgaard, Samuel Jansson, Don R Reynolds and Sam Cook for helpful discussions and comments.

Author contributions
K.R. wrote the first draft, produced figures, and conducted data analysis. K.R., E.B., and L.S. developed paper outline and structure. E.B. contributed to the introduction, field validation, discussion, and conclusion. L.S. contributed to the data processing section and discussion. K.P. and L.M. contributed to the instrument software development. A.S., R.L., and F.E. contributed to the instrument development and instrument characterization section. M.S. contributed with editing and contributed to figures. S.H., B.B., C.G., and J.L. collected and counted insects during the field trials. T.N. led the development of the instrumentation. J.L. took over development leadership in 2020.

Funding
This work was supported by Innovation Fund Denmark under Grant nos. 9078-00183B and 9066-00051B (9065-00239B, 9066-00051A), the Danish Environmental Protection Agency under Grant no. MST-667-00253 and Norsk Elektro Optikk AS.

Competing interests
All authors are or were (partly) affiliated with FaunaPhotonics, the company that developed the sensor described in this study.

Additional information
Correspondence and requests for materials should be addressed to K.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Paper III: Dynamic β-VAE for quantifying biodiversity by clustering optically recorded insect signals

Ecological Informatics 66 (2021) 101456

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Dynamic β-VAEs for quantifying biodiversity by clustering optically recorded insect signals

Klas Rydhmer [a,b,*], Raghavendra Selvan [c,d]

[a] Department of Geosciences & Natural Resource Management, University of Copenhagen, Copenhagen, Denmark
[b] FaunaPhotonics A/S, Copenhagen, Denmark
[c] Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
[d] Department of Neuroscience, University of Copenhagen, Copenhagen, Denmark

ARTICLE INFO

Keywords:
Unsupervised clustering
VAE
Insect classification
Biodiversity

ABSTRACT

While insects are the largest and most diverse group of terrestrial animals, constituting ca. 80% of all known species, they are difficult to study due to their small size and similarity between species. Conventional monitoring techniques depend on time consuming trapping methods and tedious microscope-based work by skilled experts in order to identify the caught insect specimen at species, or even family level. Researchers and policy makers are in urgent need of a scalable monitoring tool in order to conserve biodiversity and secure human food production due to the rapid decline in insect numbers.

Novel automated optical monitoring equipment can record tens of thousands of insect observations in a single day and the ability to identify key targets at species level can be a vital tool for entomologists, biologists and agronomists. Recent work has aimed for a broader analysis using unsupervised clustering as a proxy for conventional biodiversity measures, such as species richness and species evenness, without actually identifying the species of the detected target.

In order to improve upon existing insect clustering methods, we propose an adaptive variant of the variational autoencoder (VAE) which is capable of clustering data by phylogenetic groups. The proposed *dynamic* β-VAE dynamically adapts the scaling of the reconstruction and regularization loss terms (β value) yielding useful latent representations of the input data. We demonstrate the usefulness of the dynamic β-VAE on optically recorded insect signals from regions of southern Scandinavia to cluster unlabelled targets into possible species. We also demonstrate improved clustering performance in a semi-supervised setting using a small subset of labelled data. These experimental results, in both unsupervised- and semi-supervised settings, with the dynamic β-VAE are promising and, in the near future, can be deployed to monitor insects and conserve the rapidly declining insect biodiversity.

1. Introduction

Insects make up the majority of all known animal species with ca. 1 million described species and an estimated 3–4 million yet to be discovered (May, 1988; Stork, 2018). While insects are numerous and found in almost all habitats, the total insect population is thought to be shrinking at an alarming rate. An influential report recently reported a 70% loss of flying insect biomass in 30 years (Hallmann et al., 2017). These losses have mainly been driven by changes in the agricultural landscape, increased use of pesticides and the spread of disease, but the exact reasons and consequences are still unknown (Potts et al., 2016; Goulson et al., 2015). In order to accurately measure the biodiversity and health of the insect community across various biotopes (or habitats), researchers, agronomists, policy makers and institutions are in need of insect monitoring capabilities from multiple areas, over long periods of time.

Conventional insect biodiversity monitoring typically involves various trapping methods, each with their own bias towards different species, which makes it difficult to compare the results across studies (Muirhead-Thompson, 2012). The collected insect specimens are further identified under microscopes by highly trained experts. These methods provide data with very high specificity but are time consuming and expensive which severely limits the ability to collect data on a large scale, or over extended time periods, with high spatial and temporal

* Corresponding author at: Department of Geosciences & Natural Resource Management, University of Copenhagen, Copenhagen, Denmark.
E-mail address: kkgr@ign.ku.dk (K. Rydhmer).

https://doi.org/10.1016/j.ecoinf.2021.101456
Received 7 July 2021; Received in revised form 4 October 2021; Accepted 4 October 2021
Available online 12 October 2021

resolution.

In recent years, new technologies have been developed for insect monitoring such as automated traps (Potamitis et al., 2014, 2017), acoustic methods (Ganchev et al., 2007; Mankin et al., 2011) and optical instruments such as the entomological lidar (Brydegaard, 2014; Genoud et al., 2018; Jansson and Brydegaard, 2018; Shaw et al., 2005). In general, these methods provide large amounts of data with a high temporal resolution but with lower specificity compared to conventional methods (Kirkeby et al., 2021; Fanioudakis et al., 2018; Potamitis et al., 2017; Chen et al., 2014). The introduction of automated and continuous monitoring methods has the potential to greatly improve biodiversity monitoring and, consequently, conservation efforts. In order to utilize the full potential of these new methods, large number of unlabelled insect recordings need to be translated into a quantifiable biodiversity index, comparable to conventional estimates.

In this work, we combine the rich data from lidar entomology with the powerful capabilities of variational auto-encoders (VAEs) (Kingma and Welling, 2014). A common trade-off that is difficult to achieve in VAEs is between its two loss components: reconstruction loss and regularization loss. Balancing these two components can yield useful low-dimensional features (representations) of the input data which can be further analyzed to perform clustering of the input data (Bengio et al., 2013; Higgins et al., 2016). We introduce a dynamically changing formulation of the scaling of the loss terms (β). The proposed β dynamics takes instantaneous changes to the loss terms and the historical performance during training into account to keep both the reconstruction and regularization performance near their optimum.

The proposed *dynamic* β-VAE is trained on unlabelled insect recordings collected using a novel optical insect sensor at several sites in southern Scandinavia. We demonstrate its ability to successfully extract features from input data into the regularized latent space, and to cluster the data into an appropriate number of clusters. We experimentally validate improvements compared to more conventional methods such as the hierarchical clustering algorithm (HCA) (Brydegaard et al., 2020; Kouakou et al., 2020) and principal component analysis (PCA). Additionally, we show that the introduction of a semi-supervised data set further improves the clustering performance on the unlabelled data when evaluated on known phylogenetic groups.

2. Background & related work

2.1. Lidar-entomology & clustering

Lidar entomology is an insect monitoring method where insects are recorded as they enter an infrared laser beam which can sometimes extend for kilometers. It might be the fastest way to record large amounts of insect data, yielding up to several tens of thousands of optically recorded insect signals per day (Brydegaard et al., 2020). This data consists of time series, where the signal intensity varies with the insect cross section and wing beats. As the wing-beat frequency (WBF) varies between insects groups, it can to some degree be used to distinguish between species, alone or with other extracted features (Kirkeby et al., 2021; Fanioudakis et al., 2018; Potamitis et al., 2017; Chen et al., 2014; Gebru et al., 2018).

While the ability to identify a number of key species from automated sensors would be greatly beneficial to the entomological community, it is not sufficient to quantify the biodiversity. Instead, the total number of species (species richness) and their relative distributions (species evenness) are the commonly used measurements. Previous efforts to derive these numbers from a large number of optically recorded insect signals have been made using HCA on the WBF power spectra (Brydegaard et al., 2020; Kouakou et al., 2020). However, in order to cover a broad range of frequencies with sufficient resolution, a high dimensional feature space is required and the distance measures are non-trivial. For reduced model complexity and improved computational and clustering performance, a reduction of the parameter space is desired.

In order to reduce the parameter space while retaining the necessary information, various algorithms for extracting the WBF and other physical properties from insect recordings have been proposed and used (Kirkeby et al., 2021; Gebru et al., 2018; Qi et al., 2015; Jansson et al., 2018; Li et al., 2020). Machine learning based methods for feature extraction, such as auto-encoders (AE), have also been used to extract additional features (Qi et al., 2015), and very recently, to cluster acoustically recorded bird songs (Rowe et al., 2021). While an AE is able to generate high quality features for classification, a known behaviour of AE is the irregularity of the latent feature space where two similar data inputs might end up with very different latent representations. This makes the extracted features from an auto-encoder unsuitable for clustering recordings of similar insect species and quantifying the diversity of the recorded insects.

2.2. VAEs and β-annealing

Variational autoencoders (VAEs) consist of a regularized probabilistic encoder-decoder pair and are some of the most powerful representation learning methods (Bengio et al., 2013; Kingma and Welling, 2014). They have seen broad applications in generative modelling and unsupervised learning tasks.

Given unlabelled input data consisting of N samples with F features, $\mathrm{x} \in \mathbb{R}^{N\times F}$, the probabilistic encoder of a VAE maps the input to the posterior density $p(\mathbf{z}|\mathbf{x})$ over the latent variable, $\mathrm{z} \in \mathbb{R}^{N\times L}$. In practice, $L << N$ and the encoder neural network approximates the true posterior density, $p(\mathbf{z}|\mathbf{x})$, with a multivariate Gaussian, $q_\theta(\mathrm{z}|\mathrm{x}) \sim \mathcal{N}(\mu_\theta, \sigma^2_\theta)$. The decoder of a VAE reconstructs the input data from the latent variable and is given by the density function $p_\varphi(\mathbf{x}|\mathbf{z})$. The encoder and decoder neural networks are parameterised by θ and φ, respectively. The optimization objective of a VAE consists of two competing terms and it can be shown to be (Kingma and Welling, 2014)

$$\mathcal{L}_{\mathrm{VAE}} = -\mathbb{E}_{q_\theta}[logp_\varphi(\mathrm{x}|\mathrm{z})] + \mathrm{KL}[q_\theta(\mathrm{z}|\mathrm{x})||p(\mathrm{z})] \tag{1}$$

$$\mathcal{L}_{\mathrm{VAE}} \triangleq \mathcal{L}_{\mathrm{rec}} + \mathcal{L}_{\mathrm{reg}} \tag{2}$$

The quality of the auto-encoded reconstructions is controlled by the reconstruction loss $\mathcal{L}_{\mathrm{rec}}$, which is the first term in Eq. (1). The encoder density is regularized to match the prior over the latent variable, $p(\mathrm{z}) \sim \mathcal{N}(0\mathrm{bf},\mathrm{I})$, enforced by the regularization loss, $\mathcal{L}_{\mathrm{reg}}$, which is the Kullback-Leibler divergence (KLD) term in Eq. (1). At a high level, the regularization term controls the smoothness or the regularity of the latent space. Well structured and smooth latent spaces can yield useful representations of the input data.

The trade-off between the two loss terms can have influence on the performance of any VAE. A VAE where the reconstruction term dominates might be able to reconstruct the input data well with a latent space that might not be interesting for the downstream tasks (such as clustering). To alleviate this, a simple trick of scaling the regularization term $\mathcal{L}_{\mathrm{rec}}$ was used in Higgins et al. (2016) resulting in a modified objective:

$$\mathcal{L}_{\beta-\mathrm{VAE}} = \mathcal{L}_{\mathrm{rec}} + \beta\mathcal{L}_{\mathrm{reg}}. \tag{3}$$

Here the role of $\beta \geq 0$ is to balance the reconstruction- and regularization losses. Typically, lower β values yield better reconstructions but a less regularized latent space and less disentangled features. On the other hand, higher β may lead to posterior collapse, where all reconstructions are reduced to the average input and the KLD approaches zero. Various methods have been proposed to overcome this instability in achieving a reasonable trade-off between the loss terms. A common implementation is β-annealing, where β is gradually increased from a very low value up to a fixed point. While this solves the initial stability problems, the task of finding the optimal value of β remains. Recently, it has been shown that repeating the process with a cyclic β can lead to better performance (Fu et al., 2019). However, when unchecked both implementations face the risk of posterior collapse (vanishing KLD) once β enters a stationary

phase.

More recently, several approaches have attempted to *adapt* β instead of using a fixed or scheduled scaling (Shao et al., 2020, 2020; Asperti and Trentin, 2020). In the controlVAE formulation in Shao et al. (2020), rather than gradually increasing β to a maximum value (annealing), it is controlled with feedback from a non-linear proportional-integral (PI) controller to keep the KLD at a desired level. This addresses the vanishing KLD problem but the users still have to set the desired value of the KLD, which might not be straightforward for many applications.

3. Methods

The primary goal of this work is to obtain low-dimensional latent representations suitable for clustering of the high-dimensional input data. Specifically, the objective is to obtain latent space encodings of the insect spectral data such that similar species of insects are positioned close to each other. To this end, we propose to dynamically adapt, the otherwise constant scaling factor, β of a standard β-VAE.

In our proposed dynamic β-VAE formulation, the changes in reconstruction and regularisation losses are monitored throughout the training process; these changes are used to adapt the β value in each epoch using a simple control algorithm. If either the reconstruction- or regularization losses increase above a specific level of their historical minimum then β is adjusted (either by increasing or decreasing) until a new optimum is attained. This dynamic control of β maintains a steady trade-off between the two loss terms while reducing the global loss function by latching on to the historical minimum of the loss components.

In the remainder of this section, we detail the dynamic β-VAE and formulate a semi-supervised variant of the model using a new clustering loss component.

3.1. Dynamic β-VAE

The key contribution in this work is an online, adaptive formulation of the β-VAE using dynamic control of β. This is achieved by varying β at each epoch, based on the instantaneous changes in the reconstruction- and regularization terms in Eq. (3), with an objective of not letting either of the loss terms to dominate the overall model optimization. This results in a trade-off between sufficiently good reconstructions and adequately regularized latent space yielding representations of the input data that are useful for the downstream task.

At any given epoch t the objective for the dynamic β-VAE is given by,

$$\mathcal{L}^{(t)} = \mathcal{L}^{(t)}_{\text{rec}} + \beta^{(t)}\mathcal{L}^{(t)}_{\text{reg}}. \tag{4}$$

The dynamically controlled $\beta^{(t)}$ is formulated using the signum function[1], $\psi[\cdot]$, given by

$$\begin{aligned}\beta^{(t)} = \beta^{(t-1)} \quad & -\frac{b}{4}\left(1-\psi\left[\Delta_{\text{reg}}\right]\right)\left(1+\psi[\Delta_{\text{rec}}]+\Delta\mathcal{L}_{\text{rec}}\right) \\ & +\frac{a}{4}\left(1-\psi[\Delta_{\text{rec}}]\right)\left(1+\psi\left[\Delta_{\text{reg}}\right]-\Delta\mathcal{L}_{\text{rec}}\right)\end{aligned} \tag{5}$$

where

$$\Delta_{\text{rec}} = \mathcal{L}^{(t)}_{\text{rec}} - w_1 min\left(\mathcal{L}^{(:t-1)}_{\text{rec}}\right) \tag{6}$$

$$\Delta_{\text{reg}} = \mathcal{L}^{(t)}_{\text{reg}} - w_2 min\left(\mathcal{L}^{(:t-1)}_{\text{reg}}\right) \tag{7}$$

[1] The signum function, $\psi[x]$ returns the sign of any real number $s \in \mathbb{R}$

$$\psi[s] = \begin{cases} +1 & s > 0 \\ 0 & s = 0 \\ -1 & s < 0 \end{cases}$$

$$\Delta\mathcal{L}_{\text{rec}} = \psi\left[\mathcal{L}^{(t)}_{\text{rec}} - w_3\mathcal{L}^{(t')}_{\text{rec}}\right] + \psi\left[\mathcal{L}^{(t)}_{\text{rec}} - w_4\mathcal{L}^{(t')}_{\text{rec}}\right] \tag{8}$$

with hyperparameters $[a, b, w_1, w_2, w_3, w_4] \in \mathbb{R}^+$. The notation $(:t-1)$ is used to indicate all epochs up to (t-1) and (t') is the epoch when β was last changed. The terms associated with $(:t-1)$ provide a form of long term memory of the previous local optima for each of the two loss terms.

The β dynamics in Eq. (5) can be divided into two regimes aimed at optimizing reconstruction- and regularization terms corresponding to increase- and decrease of β, respectively.

3.1.1. Reconstruction regime ($\beta\downarrow$)

The value of β is decreased due to the $-b$ term in Eq. (5) when Δ_{rec} is positive; meaning the reconstruction loss is increasing compared to the historical minimum reconstruction loss, according to Eq. (6). The β decrease rule also checks if the regularization loss is decreasing compared to the historical minimum with the term $1-\psi[\Delta_{\text{rec}}]$ in Eq. (5).

3.1.2. Regularization regime ($\beta\uparrow$)

The increase in β happens due to the $+a$ term in Eq. (5) when Δ_{reg} is positive; meaning the regularization loss is increasing according to Eq. (7). The increase rule checks if the reconstruction loss has decreased compared to the historical minimum with the term $1-\psi\left[\Delta_{\text{reg}}\right]$ in Eq. (5).

Additionally, $\Delta\mathcal{L}_{\text{rec}}$ in Eq (8) nudges a change in β based on the last update to β. This allows β to get out of plateaus of either stable reconstruction or regularization regimes.

In Fig. 1, one instance of optimizing the dynamic β-VAE with the objective in Eq. (4) is shown. The value of β increases until about epoch 700 at which it plateaus and decreases (Fig. 1, row 3). At epoch 2000, it has finally stabilized. These changes are correlated with changes to $\mathcal{L}_{\text{reg}}$ and $\mathcal{L}_{\text{rec}}$ captured in the second row of Fig. 1, estimated according to Eqs. (5)–(8).

3.2. Semi-supervised clustering

Using a small subset of labelled data that optimizes a relevant loss could steer learning of representations that are more expressive for the downstream tasks under consideration. One approach to achieve this is to introduce auxiliary tasks based on the labelled data, resulting in a semi-supervised learning setup (Figueroa and Rivera, 2017).

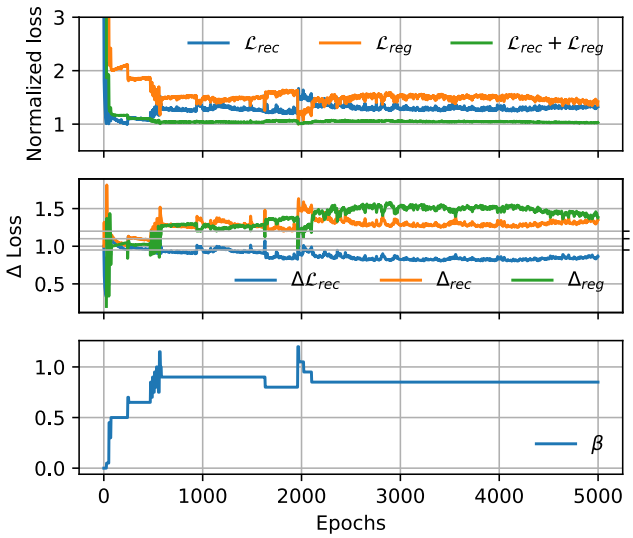


Fig. 1. The evolution of β during training of an unsupervised dynamic β VAE. After a 25 epoch warm-up phase when $\beta = 0$, it is dynamically adjusted based on $\Delta\mathcal{L}_{\text{rec}}$, Δ_{rec} and Δ_{reg}. Thereby, $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{reg}}$ remain balanced without increasing the total loss (Row 1) implying stable model convergence.

As we are interested in clustering of insect species based on their latent representation, we enforce clustering of a small subset of labelled examples to improve the overall clustering. An additional loss term, $\mathcal{L}_{\text{cls}}$, based on the auxiliary task is introduced to the model optimization:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \beta^{(t)}\mathcal{L}_{\text{reg}} + \gamma^{(t)}\mathcal{L}_{\text{cls}} \tag{9}$$

where $\gamma^{(t)}$ is the scaling of the clustering loss component.

The clustering loss, $\mathcal{L}_{\text{cls}}$, has two components to encourage intra-class cohesion and inter-class repulsion. Intra-class cohesion is captured as the sum of distance between all data points belonging to a particular class and the corresponding cluster centroid location. Assuming K clusters with centroids $\hat{\mathbf{z}}_k \in \mathbb{R}^L, k = 1, \ldots, K$ and N_k cluster members with class label k denoted $\mathbf{z}^k$, the intra-class cohesion distance is given by:

$$d_C = \sum_{k=1}^{K}\sum_{i=1}^{N_k} d(\mathbf{z}_i^k, \hat{\mathbf{z}}_k). \tag{10}$$

The inter-class repulsion distance is captured using the sum of all pairwise distances between the cluster centroids:

$$d_R = \sum_{i=1}^{K}\sum_{j<i} d(\hat{\mathbf{z}}_i, \hat{\mathbf{z}}_j) \tag{11}$$

In both cases, $d(\cdot)$ is the Euclidean distance.

Finally, the clustering loss is computed as the ratio between the two distances, which when minimized encourages smaller intra-class and larger inter-class distances, given by

$$\mathcal{L}_{\text{cls}} = \frac{d_C + \varepsilon}{d_R + \varepsilon}, \tag{12}$$

where ε is a constant used for numerical stability.

3.3. Training of dynamic β-VAE

The final objective of the dynamic β-VAE with semi-supervision in Eq. (9) has three components which are introduced during the training in three successive stages:

1. Warm-up phase ($\beta = 0$, $\gamma = 0$): In this phase, the model primarily learns to reconstruct the input data similar to bottleneck autoencoders.
2. Regularized phase ($\beta > 0$, $\gamma = 0$): In this phase, the dynamic control of β sets in which smooths the learned latent space without deterioration in the quality of reconstructions.
3. Semi-supervised phase ($\beta > 0$, $\gamma > 0$): In this phase, the learned latent space is steered to favour the downstream task of clustering.

4. Experiments & results

The main objective of the proposed dynamic β-VAE is to cluster unlabelled insect spectra into plausible clusters that could correspond to unique species. To evaluate the performance of the model, we use real data collected from field instruments and compare the model's performance in different settings. Details of the data and experiments are presented in this section.

4.1. Data collection & pre-processing

The insect data were recorded with a novel instrument from FaunaPhotonics, which uses a similar principle at close- and long ranges as the methods described in (Kirkeby et al., 2021; Gebru et al., 2018; Brydegaard et al., 2020; Brydegaard, 2014). In the current implementation, an air volume is illuminated with infrared light in two spectral bands at 808 nm and 975 nm. The back-scatter of any object passing through a 20 L volume within 1m of the sensor is recorded onto a photo diode quadrant detector. As insects fly past, the optical cross section varies with their WBF. This yields a modulated time series, sampled at 20 kHz with a bandwidth from 0 to 5 kHz. As signals from any non-insect object passing through the volume are also recorded, a CNN trained with manual labels was used to filter out insect recordings from rain and dust etc. For simplicity, the multi spectral time series were reduced to one dimension by calculating the average Welch power spectra (Welch, 1967) over both spectral bands in $F = 193$ bins between 0 and 2 kHz. Finally, the data was log-transformed and individually normalized by the maximum of each spectrum.

The unlabelled data were recorded from March to November 2020 in various biotopes in the Öresund region in southern Scandinavia and $N = 40,000$ insect recordings were randomly selected for the experiments with $F = 193$ features after the WBF pre-processing.

Additionally, data encompassing 12 different species groups were labelled one species at a time in closed cages in Copenhagen, Denmark. From this data, 6000 insect recordings (15% of the unlabelled data) were randomly selected and added to the labelled training set. For each species group, this resulted in 500 labelled recordings to be used in the semi-supervised mode. The average WBF spectra for each labelled species group is shown in Fig. 2.

In order to validate the clustering ability of the different models, 8 out of the 12 labelled species were included in computing the clustering loss, $\mathcal{L}_{\text{cls}}$ in Eq. (11), in the semi-supervised setting. The remaining four labelled species were used as test set for validating the clustering accuracy.

4.2. Experimental set-up

The dynamic β-VAE was evaluated in unsupervised and semi-supervised modes to obtain latent representations, which were clustered using K-means (Lloyd, 1982). Their clustering performance was compared with the baseline methods: PCA, Kernel-PCA, HCA using the standard implementations in sklearn (Pedregosa et al., 2011) and a conventional VAE on the same data. The encoder neural network $q_\theta(\mathbf{z}|\mathbf{x})$ consists of 9 fully connected layers, with rectified linear unit (ReLU) activation (except for the last layer). The encoder predicts the mean and the variance of the approximate posterior distribution. The decoder neural network $p_\varphi(\mathbf{x}|\mathbf{z})$ is implemented with 10 fully connected layers and ReLU activation (except the last layer, which has sigmoid activation). The VAE uses a bottleneck $L = 2$ to create the latent representation. The model layout was developed on a independent unlabelled dataset recorded at a different location and was gradually expanded until reconstructions were sufficiently good. In order to visualize the latent representation, the size of the bottleneck of the model (latent dimension) was limited to two. Details of the network architecture are reported in Table 1.

Both the unsupervised and semi-supervised models were run five times on random training and test splits. A random subset of 3000 recordings from the dataset were used as the test set.

After each training run the latent representation of the unlabelled test set was clustered using K-means method. The appropriate number of clusters K were automatically selected by the maximum average silhouette score (Rousseeuw, 1987) from a range of $5 - 50$. As we expect the unlabelled data to consist of at least 5 distinct species, we incorporate this as prior information in choosing the range of clusters. For comparison, the data were also clustered into the same range of clusters using PCA, Kernel-PCA (implemented with sigmoid kernels) and HCA (implemented with complete linkage).

The final evaluation was done by comparing the automatically identified clusters with the four labelled test species. The automatically found clusters were compared with the labelled data using the adjusted rand index (ARI) (Rand, 1971) and adjusted mutual (AMI) information score (Vinh et al., 2010). These metrics are useful to compare clusters obtained in unsupervised settings, as they are agnostic to labels and only focus on the similarity between members within the clusters.

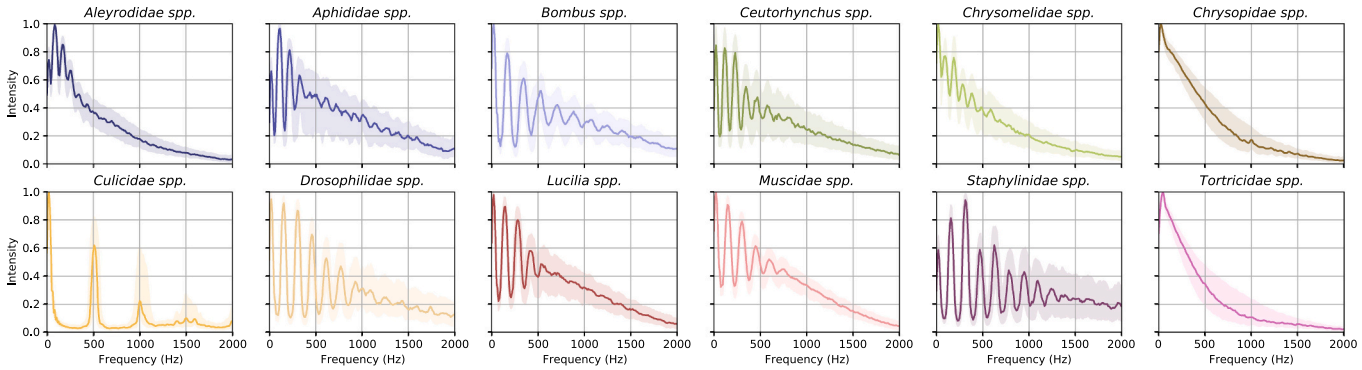


Fig. 2. The median wing-beat frequency (WBF) spectra estimated from each labelled species in the input data, $\mathbf{x}_i \in \mathbb{R}^F$. The shaded areas indicate the inter-quartile range, between 25% and 75%. The mosquitoes (*Culicidae* spp.) have the highest WBF and the moths (*Tortricidae* spp. the lowest. The weevils, (*Ceutorhynchus* spp.) have a large variation around their fundamental WBF. All recordings are log-transformed and individually normalized.

Table 1
Network architecture of the implemented dynamic β-VAE showing the number of hidden units per layer (H) and the non-linear activation functions per layer in the encoder and decoder parts of the network. (RL: Rectified Linear Unit. SG: Sigmoid.)

#		1	2	3	4	5	6	7	8	9	10
Enc.	H	193	128	128	64	32	16	8	4	2 + 2	–
	Act.	RL	RL	RL	RL	RL	RL	RL	RL	–	–
Dec.	H	2	4	8	16	32	32	64	128	128	193
	Act.	RL	RL	RL	RL	RL	RL	RL	RL	RL	SG

4.3. Model parameters & hyperparameters

The dynamic β-VAE has several tunable model parameters, as seen in Eqs. (3)–(8). These model parameters were tuned on an independent dataset, collected with the identical instrumentation but at a different location. The data had a similar distribution as the data used in this work and we obtained: $a = 0.2$ and $b = 0.05$, $w_1 = w_2 = 1.2$, $w_3 = 0.9$ and $w_4 = 1.1$. These parameters were found to be sufficiently robust on the dataset used in this work without any fine-tuning.

All models were implemented in PyTorch (Paszke et al., 2019) and trained for 5000 epochs using the Adam optimizer (Kingma et al., 2015) with a learning rate of 10^{-3}. The models were trained on Nvidia GTX 1050 graphics processing unit with 4 GB memory with a batch size of 256. A decision to adapt β was taken every fifth epoch to avoid random fluctuations. The scaling of the clustering loss, γ, in the semi-supervised mode was cycled between 0.01 and 0.2 every 100 epochs.

4.4. Results

The clustering performance on the labelled test set for the unsupervised and semi-supervised instances of the dynamic β-VAE is presented in Table 2.

The dynamic β-VAE performs better than the baselines in the ARI- and AMI-scores which quantifies the intra-class cohesion and inter-class separability. While HCA have been successfully used to identify groups of similar insects by other groups previously (Brydegaard et al., 2020), it has the lowest ARI score. With semi-supervision the dynamic β-VAE further improves upon its unsupervised clustering scores, and the improvements compared to the conventional models are more pronounced. The nonlinear kernel-PCA does not show any drastic improvements over conventional PCA. The low dimensional representations of the test species for each model are shown in Fig. 4. While the different species form largely separable and homogeneous clusters in all methods, they are relatively more compact in the semi-supervised implementation.

Table 2
Aggregated results from 5 repetitions of each method. The unsupervised model performs better than the classical models and adding the labelled data further improves clustering of unlabelled data. The median number of automatically determined clusters (K) are also reported. (ARI and AMI scores: higher is better.

Models	K	ARI-score	AMI-score
PCA	5	0.15 ± 0.02	0.21 ± 0.01
K-PCA	5	0.17 ± 0.02	0.22 ± 0.01
HCA	16	0.11 ± 0.06	0.21 ± 0.10
VAE	5	0.14 ± 0.09	0.20 ± 0.09
β-VAE	7	0.25 ± 0.02	0.34 ± 0.03
β-VAE (semi-sup.)	6	0.28 ± 0.05	0.37 ± 0.05

In the results presented in Table 2, the appropriate number of clusters found in the unlabelled test set data is also reported. The number of clusters, K, was automatically chosen to maximize the average silhouette score (Rousseeuw, 1987).

An example of the latent representation from all 12 labelled species groups by the unsupervised instance is shown in Fig. 3. All species generate dense, but partly overlapping, clusters except the weevils (*Ceutorhynchus* spp.), and to some degree the fruitflies (*Drosophilidae* spp.) which form sparser clusters.

The latent space obtained by the semi-supervised β-VAE on the unlabelled test set is shown in Fig. 6a. Using $K = 15$ the data is color coded by cluster and the average spectra from each cluster is shown in Fig. 6b.

5. Discussions

In this work, we introduced a dynamic β-VAE in order to achieve a good trade-off between the reconstruction and regularization loss terms by performing online adjustment of the β term. The proposed β dynamics result in useful latent representations for the downstream clustering task. Our experiments demonstrate the ability of the model to map high-dimensional insect data into a well regularized latent representation where phylogentic groups are distinguishable.

5.1. Generalization of the β dynamics

The primary objective of using the β dynamics in Eq. (4) is to perform *online* adjustment of the scales of reconstruction and regularization terms based on their instantaneous values while taking the previous

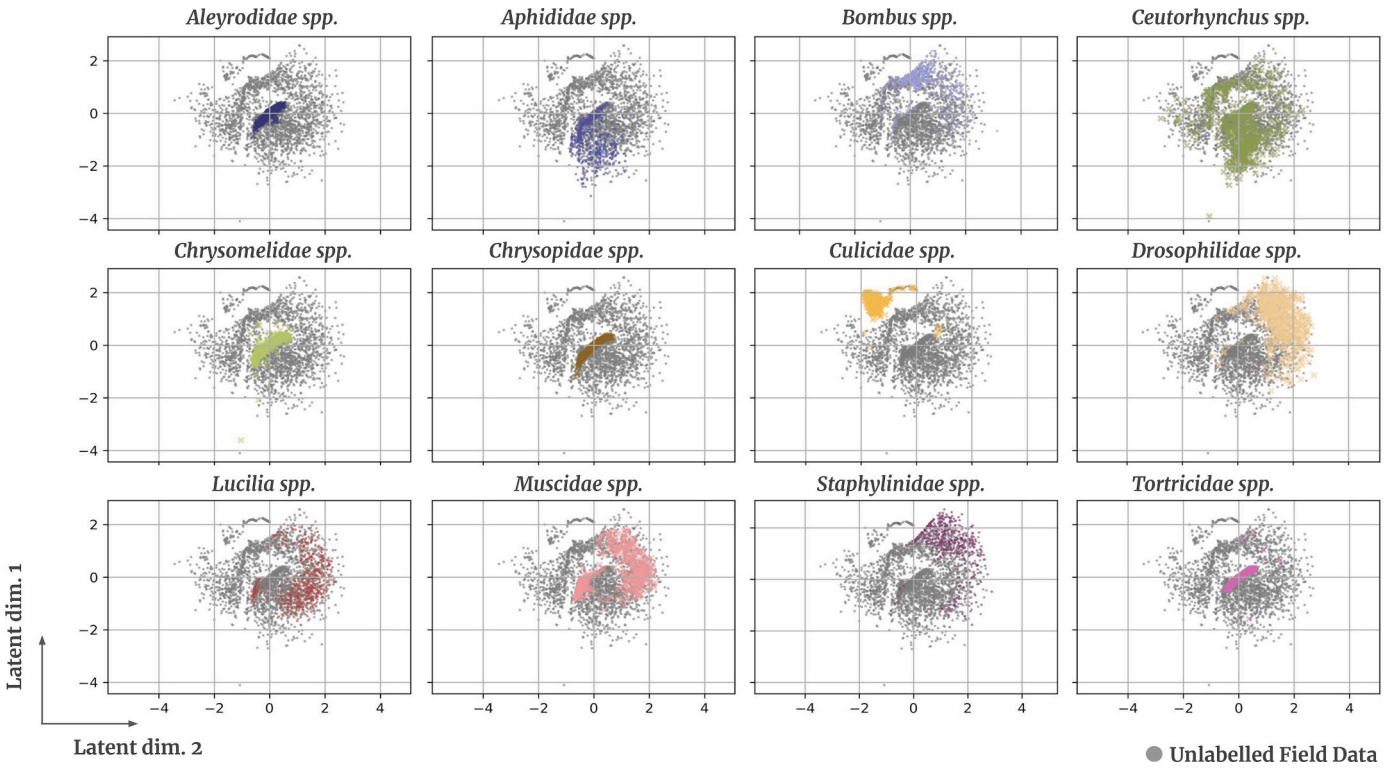


Fig. 3. Latent representation of optically recorded insect wingbeat frequency spectra. The proposed dynamic β-VAE is able to cluster unlabelled field recordings into compact clusters. Evaluated on labelled data, most species form compact clusters, as shown with different colours for each of the 12 named species groups. Closely related species groups, such as the *Lucilas*pp. and *Muscidae* spp. are partially overlapping but well separated from more distant species groups, such as the *Bombus* spp.

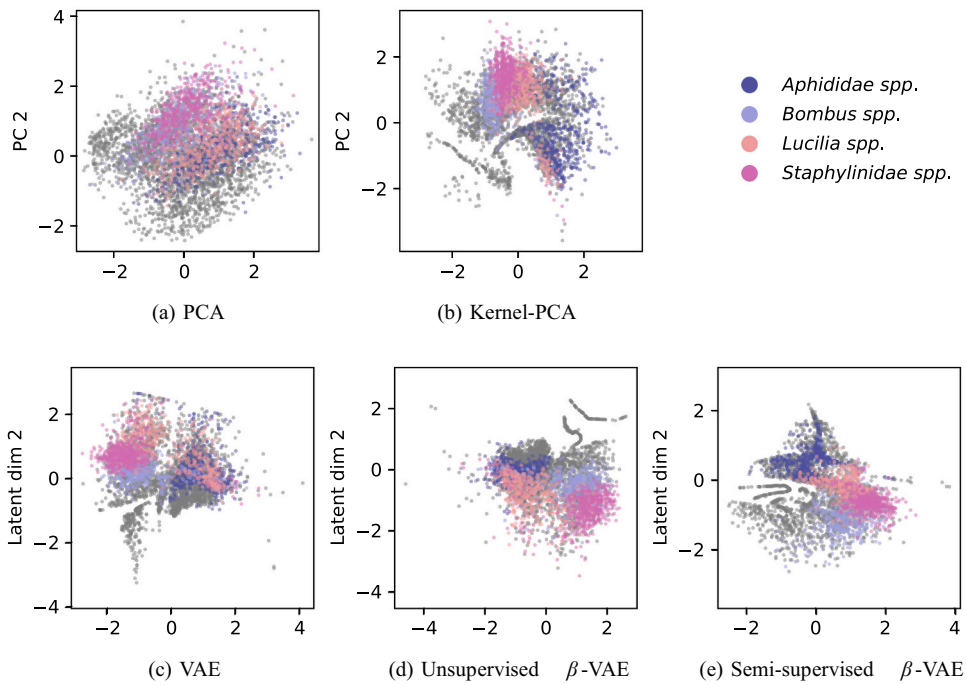


Fig. 4. Example latent representation of unlabelled field data and the four test species. While all methods map the high dimensional input data to a two dimensional feature space, the dynamic β-VAEs creates clusters with less overlap between the species groups. The inclusion of 10% labelled data for training further improves the results, yielding denser clusters with less overlap than the unsupervised β-VAE.

optima into account. The specific formulation of these control mechanisms in Eqs. (6)–(8) force the model optimization to not deviate from the previous optimal solutions. The terms comprising *min* over $(: t-1)$ epochs in Eqs. (6) and (7) provide a form of memory of the previous local optima. The trade off between long and short term memory of the losses and the corresponding optima help the model to steer towards more global optima. These equations provide a sufficiently general formulation for adjusting β as they are only dependent on the two loss components. Further, one can also envision a learnable neural network with long short-term memory (LSTM) that can perform this dynamic control in a recurrent neural network type formulation of a closed loop control system (Hochreiter and Schmidhuber, 1997).

5.2. Influence of the β and γ parameters

As seen in Fig. 1, the initial effect of a dynamic β is similar to β-annealing, where β gradually increases during training in order to prevent posterior collapse (Bowman et al., 2015). However, the key difference with β annealing is that the rate of annealing is not predetermined, as the β dynamics described in Section 3.1 enables self-regulation of β. This is witnessed during the latter part of training, where β repeatedly adapted either by increasing or decreasing its value dependent on the changes in the reconstruction- and regularization terms. This behaviour is similar to what is reported with the cyclic β-VAE (Fu et al., 2019) where the *shakeup* often allows the model to obtain new global minimum loss. The similarity to the cyclic β-VAE is further enhanced by automatically increasing β when there has been no change for a large number of epochs (500 in our case). However, unlike a cyclic β-VAE, β is only cycled if it has reached a stationary condition and Δ_{reg} and Δ_{rec} are within limits. This helps both the unsupervised- and semi-supervised instances of dynamic β-VAE to latch on the historical minimum of both $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{reg}}$.

The β dynamics introduced in Section 3.1 is also similar to adaptive strategies used in models such as the controlVAE (Shao et al., 2020). A controlVAE stabilizes the model by adjusting β to keep the regularization loss (KLD) at a constant level. However, finding an appropriate KLD level can be difficult as it could vary across datasets and downstream tasks. In contrast, the dynamic β-VAE keeps the model stable by constantly comparing both $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{reg}}$ with their historical minima. A gain on either loss term, at the expense of the other, is counteracted by adjusting β. This self-regulating β dynamics that is not dependent on fixing KLD value is an advantage with our formulation.

Including the additional loss term scaling term $\gamma^{(t)}\mathcal{L}_{\text{cls}}$ in Eq. (9) further improved the clustering performance of the model. In this implementation γ was cycled between 0.01 and 2 in order to keep the contribution from $\mathcal{L}_{\text{cls}}$ in a similar range as $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{reg}}$. A logical next step could be to expand Eq. (4) to include a dynamically adjusted $\gamma^{(t)}\mathcal{L}_{\text{cls}}$ term. This would however make the model less generalized to other tasks.

5.3. Performance comparison

Comparing the performance between the models reported in Table 2, the dynamic β-VAE perform better than the baseline models. Adding 15% of labelled data from 8 different species to the training set further improves the clustering performance. While using HCA on high dimensional frequency spectra have been successfully used to identify mosquito clusters in field data (Brydegaard et al., 2020) and biodiversity evaluation (Kouakou et al., 2020), our results show better performance for PCA + Kmeans. While the Kernel-PCA generally produced more heterogeneous latent distributions, it did not show any significant improvements over the standard PCA.

The non adaptive β-VAE showed large variation in its performance but was on average comparable with the conventional methods. The dynamic β-VAEs kept $\beta < 1$ during most of the training, as exemplified in Fig. 1, and since a higher β term favours a well generalised latent space over good reconstructions, a reduction in clustering performance could be expected. Additionally, the non-adaptive VAE was more cumbersome to train as the model collapsed frequently during training.

5.4. Selection of number of clusters

In this work, the appropriate number of clusters was automatically selected by maximizing the average silhouette score. However, when manually evaluating the average silhouette score and comparing it with commonly used empirical measurements, such as the elbow method (Thorndike, 1953) and intra-cluster sum-of-squares, a user might typically identify a higher number of clusters. Having more clusters yield more similar recordings within each cluster. An example is show in Fig. 6a and b where the number of clusters were manually selected. As in previous work by lidar-entomologists (Brydegaard et al., 2020), some species groups can be identified from these clusters at this level by comparing the average spectra of each cluster with known data. For example a cluster of possible mosquitoes can be identified by their high wing-beat frequency the lower right corner of Fig. 6b.

With the 3000 randomly chosen insect recordings from multiple sites during summer, we expect the total number of species represented in the test set to be one or several orders of magnitude larger. We tested a range of $5-50$ clusters as even reasonably coarse clustering will be useful for quantifying biodiversity. Once fully deployed on a network of insect sensors, a dataset recorded in an environment with high biodiversity could yield more clusters than a dataset captured in a biologically poor environment. This would allow the automated and optically recorded insect data to be correlated with conventional monitoring methods and greatly improve the ability to monitor insect biodiversity at scale.

5.5. Computation time and inference

Computation time for PCA + K-Means is significantly shorter compared to HCA on the full spectra (Brydegaard et al., 2020; Kouakou et al., 2020). While the initial training of the dynamic β-VAE takes a few hours depending on the number of epochs, once trained, the inference time for the model is comparable with that of using PCA + K-Means. Once deployed in the field, the dynamic β-VAE model is not expected to be retrained regularly but to be used as a dimensionality reduction method. Therefore, inference time is a more important metric than the initial computation time.

5.6. Exploring the latent space

Samples generated from the latent space of the semi-supervised model are shown as a *latent space cart-wheel* in Fig. 5. Traversing different lines in the latent space results in samples that smoothly transition between different spectra types. As a side note, we point that the two latent features do not appear to be entirely disentangled; this is manifested as dense islands and sparse spaces of spectra in the latent space. For our downstream clustering task, fully disentangled features are not required. However, one could introduce an additional loss component that enforces orthogonality between the different latent dimensions to achieve improved disentanglement.

The latent representation of the unsupervised model can be further validated by comparing Fig. 3 with the average spectra of each group, shown in Fig. 2. Species groups with similar spectra, such as *Aleyrodidae* spp., *Aphididae* spp., *Tortricidae* spp. and *Chrysopidae* spp. are positioned in similar areas. Similarily, all dipterans (*Lucilia* spp., *Muscidae* spp. and *Drosophila* spp.) have overlapping clusters except the mosquitoes (*Culicidae* spp.) which have a much higher WBF and are more isolated. In Fig. 3, the model performs less well on the weevils (*Ceutorhynchus* spp.) compared to the other species. It is likely due to a larger variation in their WBF than for the other groups.

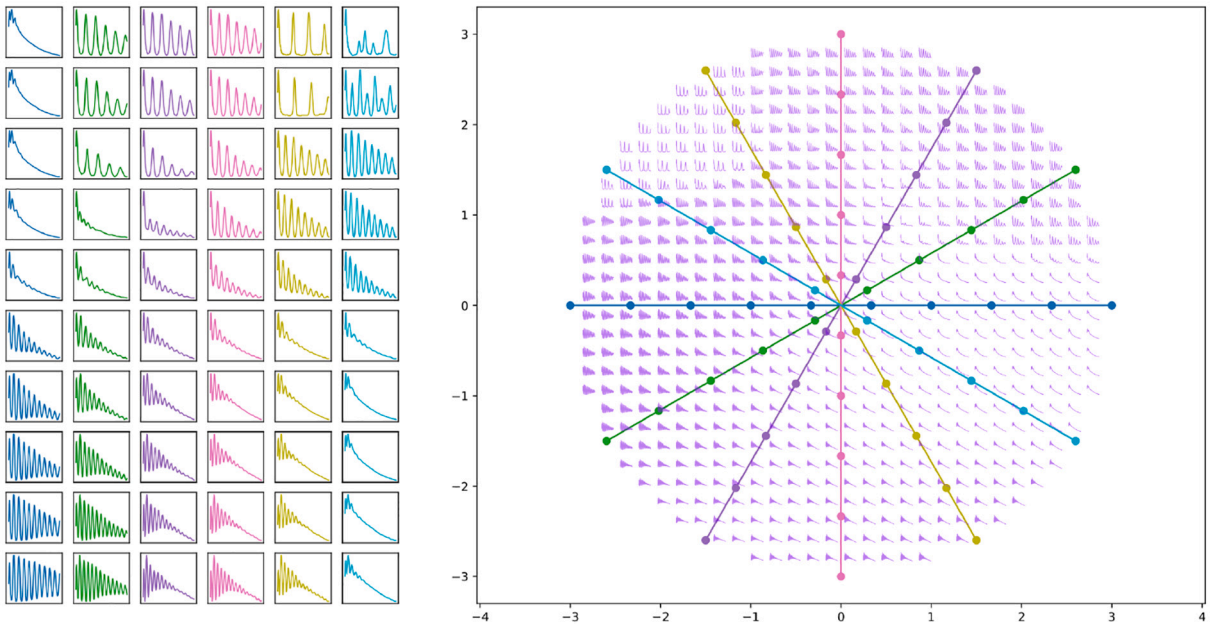
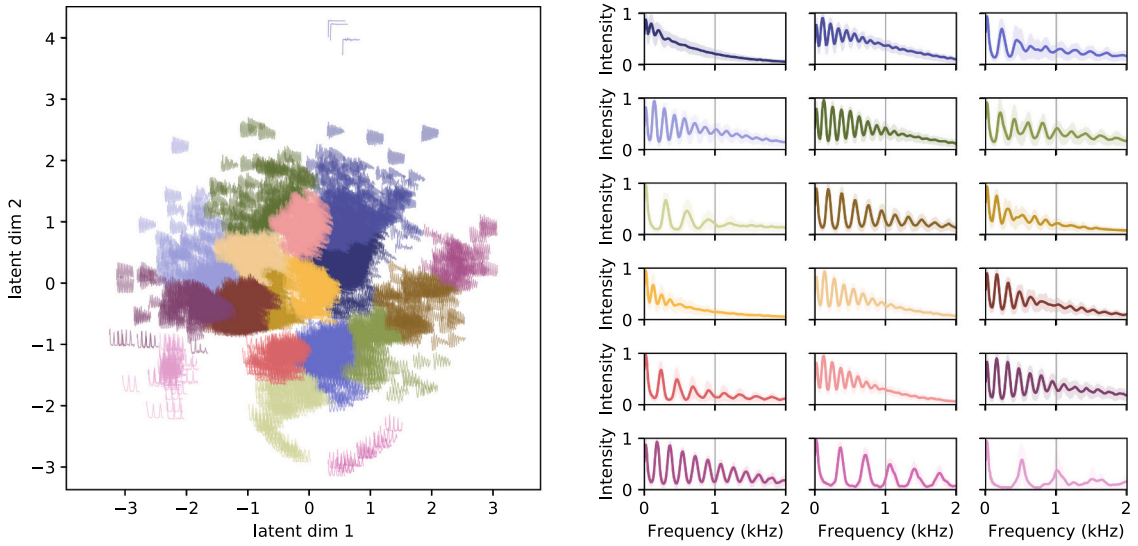


Fig. 5. Latent space cart-wheel visualization. Samples from the latent representation of the semi-supervised dynamic β-VAE are shown with decoded latent samples when the lines are traversed. While the two dimensions do not appear fully disentangled, the latent space is regularized and transitions between various areas are smooth and gradual.



(a) Latent representation (b) Average WBF spectra and IQR in each cluster

Fig. 6. K-means clustering in the latent representation of unlabelled field data from the semi-supervised β-VAE into 15 clusters. The model is capable of generating a low dimensional space where similar insect recordings are represented clustered together. The lower right cluster in (b) are likely to contain mosquitoes due to their high WBF.

6. Conclusions

In this work, we have presented, to our knowledge, the first VAE designed for clustering of optically recorded insect signals. The dynamic β-VAE was developed in order to achieve a stable model, optimizing both reconstruction- and regularization terms. When trained on unlabelled data recorded during field conditions, the model is able to automatically create meaningful clusters. The unsupervised clustering performance was validated with labelled data collected in controlled conditions showing promising results. By using 15% of labelled data during the training process for semi-supervision, this clustering performance is further improved. Already at this stage, it is possible to extract easily identifiable insect groups such as mosquitoes from the automatically identified clusters and we expect this capability to grow as the collection of labelled reference data continues.

The future aim of our work is to further improve automatic identification of the number of clusters. In the near future, the model will be deployed on several sensors in the field, and estimated cluster sizes and

distributions will be compared to conventional methods. This will greatly improve monitoring possibilities and decision support tools for entomologists and agronomists. In order to mitigate the trend of declining insect communities, the first step is to ensure adequate data collection. In the near future, we believe methods based on the proposed dynamic β-VAE will be useful to quantify and, thus, help conserve insect biodiversity.

Funding

This work was supported by the Innovation Fund Denmark.

Data availability

The source code and a sample dataset are available here: https://github.com/remhdyr/dynamicBeta

Declaration of Competing Interest

Klas Rydhmer is employed at FaunaPhotonics. We declare that this have not affected the reported results or interpretations in any way. Raghavendra Selvan declares no competing interests.

Acknowledgments

Ankit Kariyaa, Inger Kappel Schmidt and Laurence Still for proof reading. Marta Montaro and teammates for labelled data recordings. Erik B. Dam and Christoffer Grønne for helpful discussions.

References

Asperti, A., Trentin, M., 2020. Balancing reconstruction error and Kullback-Leibler divergence in variational autoencoders. IEEE Access 8, 199440–199448.
Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 35 (8), 1798–1828.
Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A.M., Jozefowicz, R., Bengio, S., 2015. Generating sentences from a continuous space arXiv preprint arXiv:1511.06349.
Brydegaard, M., Jansson, S., Malmqvist, E., Mlacha, Y.P., Gebru, A., Okumu, F., Killeen, G.F., Kirkeby, C., 2020. Lidar reveals activity anomaly of malaria vectors during pan-African eclipse. Sci. Adv. 6 (20) eaay5487.
Brydegaard, M., 2014. Advantages of shortwave infrared lidar entomology. Laser Applications to Chemical Security and Environmental Analysis. Optical Society of America. LW2D-6.
Chen, Y., Why, A., Batista, G., Mafra-Neto, A., Keogh, E., 2014. Flying insect classification with inexpensive sensors. J. Insect Behav. 27 (5), 657–677.
Fanioudakis, E., Geismar, M., Potamitis, I., 2018. Mosquito wingbeat analysis and classification using deep learning. In: 2018 26th European Signal Processing Conference (EUSIPCO). IEEE, pp. 2410–2414.
Figueroa, J.A., Rivera, A.R., 2017. Learning to cluster with auxiliary tasks: a semi-supervised approach. In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, pp. 141–148.
Fu, H., Li, C., Liu, X., Gao, J., Celikyilmaz, A., Carin, L., 2019. Cyclical annealing schedule: A simple approach to mitigating KL vanishing arXiv preprint arXiv: 1903.10145.
Ganchev, T., Potamitis, I., Fakotakis, N., 2007. Acoustic monitoring of singing insects. In: IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Vol. 4. https://doi.org/10.1109/ICASSP.2007.367014. IV-721-IV-724.
Gebru, A., Jansson, S., Ignell, R., Kirkeby, C., Prangsma, J.C., Brydegaard, M., 2018. Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight. J. Biophoton. 11 (8) e201800014.
Genoud, A., Basistyy, R., Williams, G., Thomas, B., 2018. Optical remote sensing for monitoring flying mosquitoes, gender identification and discussion on species identification. Applied Physics B 124, 1–11.
Goulson, D., Nicholls, E., Botí as, C., Rotheray, E.L., 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 347 (6229).
Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., et al., 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE 12 (10) e0185809.
Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A., 2016. β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. https://openreview.net/forum?id=Sy2fzU9gl.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neu. Comput. 9 (8), 1735–1780.
Jansson, S., Brydegaard, M., 2018. Passive KHZ lidar for the quantification of insect activity and dispersal. Anim. Biotelem. 6 (1), 1–10.
Jansson, S., Atkinson, P., Ignell, R., Brydegaard, M., 2018. First polarimetric investigation of malaria mosquitoes as lidar targets. IEEE J. Select. Topics Quant. Electr. 25 (1), 1–8.
Kingma, D.P., Welling, M., 2014. Auto-encoding variational bayes. In: International Conference on Learning Representations.
Kingma, D.P., Ba, J., Adam:, 2015. A method for stochastic optimization. In: International Conference on Learning Representation.
Kirkeby, C., Rydhmer, K., Cook, S.M., Strand, A., Torrance, M.T., Swain, J.L., Prangsma, J., Johnen, A., Jensen, M., Brydegaard, M., et al., 2021. Advances in automatic identification of flying insects using optical sensors and machine learning. Sci. Rep. 11 (1), 1–8.
Kouakou, B.K., Jansson, S., Brydegaard, M., Zoueu, J.T., 2020. Entomological scheimpflug lidar for estimating unique insect classes in-situ field test from ivory coast. OSA Continuum 3 (9), 2362–2371.
Li, M., Jansson, S., Runemark, A., Peterson, J., Kirkeby, C.T., Jönsson, A.M., Brydegaard, M., 2020. Bark beetles as lidar targets and prospects of photonic surveillance. J. Biophoton. 14 e202000420.
Lloyd, S., 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theory 28 (2), 129–137.
Mankin, R., Hagstrum, D., Smith, M., Roda, A., Kairo, M., 2011. Perspective and promise: a century of insect acoustic detection and monitoring. Am. Entomolog. 57 (1), 30–44.
May, R.M., 1988. How many species are there on earth? Science 241 (4872), 1441–1449.
Muirhead-Thompson, R., 2012. Trap Responses of Flying Insects: The Influence of Trap Design on Capture Efficiency. Academic Press.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32, 8026–8037.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Potamitis, I., Rigakis, I., Fysarakis, K., 2014. The electronic mcphail trap. Sensors 14 (12), 22285–22299.
Potamitis, I., Eliopoulos, P., Rigakis, I., 2017. Automated remote insect surveillance at a global scale and the internet of things. Robotics 6 (3), 19.
Potts, S.G., Ngo, H.T., Biesmeijer, J.C., Breeze, T.D., Dicks, L.V., Garibaldi, L.A., Hill, R., Settele, J., Vanbergen, A., 2016. Summary for policymakers of the assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. Intergovernmental SciencePolicy Platform on Biodiversity and Ecosystem Services.
Qi, Y., Cinar, G.T., Souza, V.M., Batista, G.E., Wang, Y., Principe, J.C., 2015. Effective insect recognition using a stacked autoencoder with maximum correntropy criterion. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–7.
Rand, W.M., 1971. Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. 66 (336), 846–850.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.
Rowe, B., Eichinski, P., Zhang, J., Roe, P., 2021. Acoustic auto-encoders for biodiversity assessment. Ecol. Inform. 62, 101237.
Shao, H., Lin, H., Yang, Q., Yao, S., Abdelzaher, T., 2020. Dynamicvae: Decoupling reconstruction error and disentangled representation learning arXiv preprint arXiv: 2009.06795.
Shao, H., Yao, S., Sun, D., Zhang, A., Liu, S., Liu, D., Wang, J., Abdelzaher, T., 2020. Controlvae: controllable variational autoencoder. In: International Conference on Machine Learning. PMLR, pp. 8655–8664.
Shaw, J.A., Seldomridge, N.L., Dunkle, D.L., Nugent, P.W., Spangler, L.H., Bromenshenk, J.J., Henderson, C.B., Churnside, J.H., Wilson, J.J., 2005. Polarization lidar measurements of honey bees in flight for locating land mines. Opt. Express 13 (15), 5853–5863.
Stork, N.E., 2018. How many species of insects and other terrestrial arthropods are there on earth? Ann. Rev. Entomol. 63, 31–45.
Thorndike, R.L., 1953. Who belongs in the family? Psychometrika 18 (4), 267–276.
Vinh, N.X., Epps, J., Bailey, J., 2010. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. J. Mach. Learn. Res. 11, 2837–2854.
Welch, P., 1967. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. 15 (2), 70–73.

Paper IV: Automating an insect biodiversity metric using distributed optical sensors: an evaluation across Kansas, USA cropping systems

Automating an insect biodiversity metric using distributed optical sensors: an evaluation across Kansas, USA cropping systems

Klas Rydhmer[1,2], James O. Eckberg[3], Jonathan G. Lundgren[4], Samuel Jansson[2], Laurence Still[2], John E. Quinn[5], Ralph Washington Jr.[2], Thomas Nikolajsen[2], Jesper Lemmich[2], Nikolaj Sheller[2], Alex M. Michels[4], Michael M. Bredeson[4], Steven T. Rosenzweig[3], and Emily N. Bick[2,6,7*]

[1] *Department of Geosciences and Natural Resource Management, Copenhagen University, Rolighedsvej 23, Fredriksberg C, 1958, Denmark.*
[2] *FaunaPhotonics, Støberigade 14, Copenhagen, 2450, Denmark.*
[3] *Agriculture and Food Solutions, General Mills, Minneapolis, MN 55427, United States.*
[4] *Ecdysis Foundation, 46958 188th St, Estelline, SD 57234, United States.*
[5] *Department of Biology, Furman University, 3300 Poinsett Hwy, Greenville, SC 29613, United States*
[6] *Department of Entomology, University of Wisconsin-Madison, 1630 Linden Dr, Madison, WI 53706, United States.*
[7] *Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.*

Corresponding author: *Emily N. Bick, ebick@wisc.edu

Author Contributions: Klas Rydhmer aggregated the data, performed the analysis, produced figures and tables, and co-wrote the first draft of the manuscript. James O. Eckberg planned and supervised the experiment and co-wrote the introduction, discussion, and conclusion. Emily N. Bick drafted the introduction, discussion, and conclusion, contributed to the methods, and coordinated edits to the manuscript. Samuel Jansson contributed to the data analysis, the results, and the methods sections and co-wrote the draft of the manuscript. Laurence Still assisted in planning and contributed to the data analysis. Ralph Washington Jr. and Nikolaj Sheller deployed optical sensors. Thomas Nikolajsen and Jesper Lemmich assisted with planning. Mike Bredeson, Alex Michels, and Jonathan Lundgren collected and supervised the analysis of the invertebrate inventory data. Jonathan Lundgren and Steven T. Rosenzweig contributed to the planning of the experiment. Steven Rosenzweig coordinated the recruitment of cooperating farmers and farm fields to host the experiment. John E. Quinn assisted with manuscript framing and editing.

Competing Interest Statement: Klas Rydhmer, Samuel Jansson, Laurence Still, Ralph Washington Jr., and Nikolaj Sheller are or were affiliated with FaunaPhotonics, who developed the sensor used in the study, as employees or stakeholders. Emily N. Bick was funded in part by FaunaPhotonics as part of a Postdoctoral Fellowship granted by the Danish Innovation Fund.

Classification: Biological Sciences, Biophysics, and Computational Biology

Keywords: Biodiversity, Bioinventory, Insect monitoring, Optical sensing, Biophotonics

Abstract

Global ecosystems and food supply depend on insect biodiversity for key functions such as pollination and decomposition. High-resolution, accurate data on invertebrate populations and communities across scales are critical for informing conservation efforts. However, conventional data collection methodologies for invertebrates are expensive, labor intensive, and require substantial taxonomic expertise, limiting researchers, practitioners, and policymakers. Novel optical techniques show promise for automating such data collection across scales as they operate unsupervised in remote areas. In this work, optical insect sensors were deployed in

20 agricultural fields in Kansas, USA. Measurements were compared to conventional assessments of insect diversity from sweep nets and Malaise traps. Species richness was estimated on optical insect data by applying a clustering algorithm to the optical insect sensor's signal features of wing-beat frequency and body-to-wing ratio. Species richness correlated more strongly between the optical richness estimate and each of the manual methods than between the two manual methods, suggesting sensors can be a reliable indicator of invertebrate richness. Shannon- and Simpson indices were calculated for all three methods but were largely uncorrelated including between conventional methods. Although the technology is relatively new, optical sensors that are calibrated against known communities may provide next-generation insight into the spatiotemporal dynamics of invertebrate biodiversity and their conservation.

Significance Statement

The implications of this research extend from the farm level to the regional level. Much of what scientists understand about the decline of invertebrates comes from a small number of long-term studies that can be coarse and correlational in nature. High-resolution biodiversity data sets on fields to landscapes may provide the insight needed for the successful management and accounting of biodiversity by government, industry, and communities. Such high-resolution data has potential to support global efforts and coordination of biodiversity conservation.

Introduction

Invertebrate biodiversity is fundamental to ecosystem processes, functions, and services (Yang & Gratton, 2014). Monitoring invertebrate populations and communities can inform management and policy at multiple scales. Such data are critical to agriculture operations and sustainability (Landis et al., 2008). However, invertebrate biodiversity is difficult to quantify (Geiger et al., 2016b; Shortall et al., 2009) and monitor at broad spatial and temporal scales (Sánchez-Bayo & Wyckhuys, 2019; Tilman et al., 1994) [10,11]. The difficulty is largely due to the necessity of skilled labor required for taxa identification on which biodiversity quantification relies (Wägele et al., 2022) and is both limited and prohibitively costly(Gardner et al., 2008) . Common approaches to collecting insect inventories include sweep netting as well as Malaise-, pan-, and light traps. Each method has its own bias toward certain insect groups (Montgomery et al., 2021; Morris, 1960) often resulting in the concurrent use of techniques in studies and practice (LaCanne & Lundgren, 2018a).

There is a need for new technology to monitor invertebrate biodiversity in real time for agricultural systems. Such a tool would provide data to support biodiversity-focused management at field, farm, and landscape scales (LaCanne & Lundgren, 2018a) and allow for tracking of the impact of conservation measures, or the lack thereof. Automation of systems has the potential to reduce labor, time, and costs. While many automated insect monitoring tools are available for agricultural pest monitoring (Bick et al., 2023; Preti et al., 2021; Silva et al., 2013), overall, these approaches are not suitable for assessing biodiversity as they focus on the identification of indicator species, not communities eg. (LaCanne & Lundgren, 2018b; J. G. Lundgren & Fausti, 2015). The automatic quantification of invertebrate biodiversity could improve the data available for monitoring and evaluation of conservation efforts but currently, no method exists at scale(Wägele et al., 2022) despite calls for such data and analytics to inform the assessment and management of ecosystems (Garcia et al., 2023a).

Concurrently collected real-time data on invertebrate biodiversity likely would improve our understanding of insect population changes at a regional or even global scale, filling a gap in tracking of insect change. The incorporation of ‘big data’ has been shown to help mitigate some methodological biases (Geiger et al., 2016a). One such effort is the global malaise project that is using automated taxonomic identification from traps using DNA, addressing the most labor-intensive part of this method (Krishna Krishnamurthy & Francis, 2012). It is a highly promising ‘big data’ approach; unfortunately, the method over-represents known species, has an inherent sampling bias towards flying insects, and emphasized species with large mitochondrial differences. Optical entomological methods such as lidar, where an optical signal is recorded from insects flying through a beam of emitted light, can record large numbers of insect flights without using a lure. However, it is unclear how optical sensors compares to conventional methods in measuring population and communities(Garcia et al., 2023b; Rydhmer et al., 2022).

The goal of this study is to determine if the measurement of an insect biodiversity metric can be automated with the use of optical near-infrared insect sensors. In this work, we deployed sensors (Rydhmer et al., 2021) in 20 agricultural fields across six crops in Kansas, USA. The sensors were deployed alongside Malaise traps and the sites were sampled with sweep nets. Each site was evaluated on two different occasions to capture seasonal changes. Specifically, we compared manual methods to each other and with the novel biodiversity metric utilizing unsupervised clustering of data collected by a lidar-based sensing method.

Materials and Methods

Data collection

Insect populations were monitored at 20 sites (Figure 1) in June and July of 2020 using sensors alongside manual methods (Malaise traps and sweep nets). Representative agricultural crops of central Kansas were sampled including three corn, three sorghum, six soybean, one alfalfa, two pasture, and five complex cover crops. The complex cover crops consisted of approximately eight species of annual grass and forb cover crops. An autonomous near-infrared sensor (described in (Rydhmer et al., 2021) and produced by FaunaPhotonics ApS., Copenhagen SV, Denmark) was placed ~50 m from field margin and was monitored continuously for two periods of three days in June and in July. The sensor uses light-emitting diodes to transmit infrared light (810 nm & 970 nm), creating a measurement volume between 5 and 70 L, depending on insect size (Rydhmer et al.,

2021). Insects flying in front of the sensor back-scatter light, which is recorded by a photodiode as a time signal and saved for further processing.

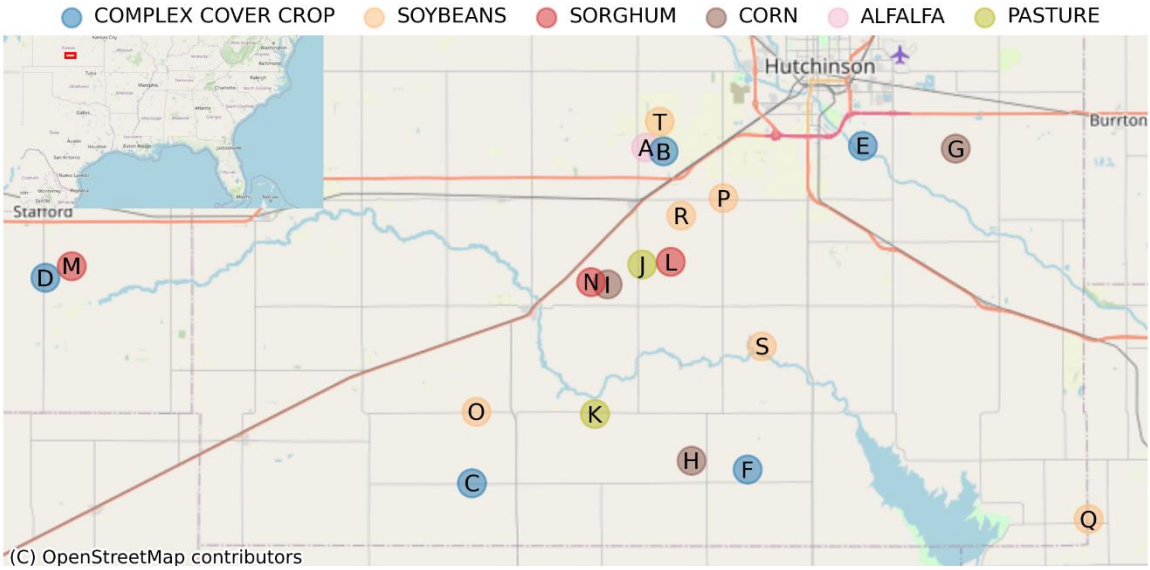


Figure 1. Map of 20 field site location distributed around central Kansas. Fields are color coded by crop type. Field A and B are separated slightly for readability. Map data from www.openstreetmap.org.

Insect recordings are automatically separated from noise originating from other sources (e.g. rain drops or plant interference) using a proprietary neural network from the instrument manufacturer, similar to [19,29]. Additionally, observations without clearly identified wingbeats or body-to-wing ratios were discarded. A total of 1,057,115 observations were recorded, of which 106,083 remained after filtering and were included in the study. A recorded observation consists of a time series data from which information pertaining to the physical features of the individual insect can be obtained (Rydhmer et al., 2021).

Sensors were compared with manual sampling of invertebrates (sweep nets and Malaise traps) in the same fields. Foliar and low flying insects were captured using a sweep net (38 cm diam., Bioquip™, Rancho Dominguez, CA, USA). Insects were collected at 50, 100, and 150 m from the field edge along a linear transect. Sweeps (n = 50 per location) were performed perpendicular to the transect, parallel to the field edge. Insects were transferred to a sealed plastic bag and were frozen until processed. In the laboratory, insects were thawed, sorted from the plant material, and identified.

Malaise traps were deployed at each site to capture the aerial insect community. A single bi-directional, Townes-style trap (dimensions 1.8 long; 1.8 m at its tallest height, and 1.2 m at its shortest height) was placed 100 m from the margin and adjacent to the ecosystem service sampling areas. The wall of the trap was parallel with the field margin. The traps were allowed to operate for 24 h, and the insects captured in the collection vials were preserved in ethanol.

All specimens collected by sweep net and malaise traps were identified to the lowest possible taxonomic unit (i.e., species or morphospecies). Due to a lack of species identification knowledge and time limitations, thrips (Insecta: Thysanoptera) were not identified beyond the family level and were not included in community metrics analyses (abundance, species richness, and diversity). All immature insects were identified to family and grouped together, except for lepidopteran larvae, which were categorized as morphospecies independent of the adult stage due to their functional differences. All other specimens were identified to species using written and online taxonomic keys. Specimens that were not able to be positively identified to species were separated into distinct morphospecies. Voucher specimens of all taxa are housed in the Mark F. Longfellow Biological Collection at Blue Dasher Farm, Estelline, SD.

Ecosystem services were evaluated for insect and weed seed predation. First, invertebrate predators were isolated from both the soil and foliar communities. Additionally, predation rates in each field were assessed using sentinel wax moths (*Galleria mellonella* L. [Lepidoptera: Pyralidae]) larvae following the (J. G. Lundgren et al., 2006) Lundgren et al., 2006protocol, using 15 sentinels per plot arranged in three 5 × 3 7.5 m grid orientations (n = 45 per field). Weed seed predation was assessed from isolating soil and foliar granivore

communities and their services using seed cards as described in Lundgren et al., 2006. Granivore services were measured on three abundant weed species (Johnsongrass (*Sorghum halapense* (L.) Pers.; Poaceae), lambsquarters (*Chenopodium album* L.; Amaranthaceae) and redroot pigweed (*Amaranthus retroflexus* L.; Amaranthaceae), V & J Seed Farms, Woodstock, IL, USA). Seeds were attached to 10 × 8 cm plastic cards (Avery™ insertable plastic dividers; #11200; Brea, CA, USA) using 6 cm strips of double-sided tape (Scotch, 3M, St Paul, MN, USA). Each species (n = 20 seeds of each species; 60 seeds per card) were placed on a 2 × 10 pattern each card. Fine quartz sand was spread over exposed areas of the tape to exclude visiting invertebrates. To exclude granivorous vertebrates, a wire cage (14 × 12 cm cage, 1.4 × 1.4 cm mesh opening) was placed over the card and placed >3 cm above it. Control cards were used to account for seed loss from environmental factors such as wind or rain and contained 1.5 × 1.5 mm black glass beads (Cousin™ DIY, #AJM61215021, Largo, FL, USA) of comparable size as the weed seeds (Lundgren et al., 2006). Each plot received three seed cards and one control card (n = 9 seed cards and three control cards per field), placed on the soil surface in the four corners of each plot. Granivory was measured as the number of seeds removed or damaged per card after a 3 d. exposure.

Data analysis

The wing-beat frequency (WBF) and body-to-wing ratio (BWR) was calculated from all observations in similar fashion to previous work by other groups(Gebru et al., 2018; Genoud et al., 2019; Kirkeby, Rydhmer, Cook, Strand, et al., 2021). The signal from the insect body (σ_b) and the diffuse and specular signal contributions were the insect wings (σ_{dw} and σ_{sw}) are estimated and separated using sliding minimum, sliding median and sliding maximum filters with a filter width corresponding to the wing beat period of the insect. The BWR is defined as the closed ratio between the body and wing contributions according to equation (1). An example of an insect signal is shown in Figure 1a.

$$BWR = \sigma_b / (\sigma_{dw} + \sigma_b) \quad (1)$$

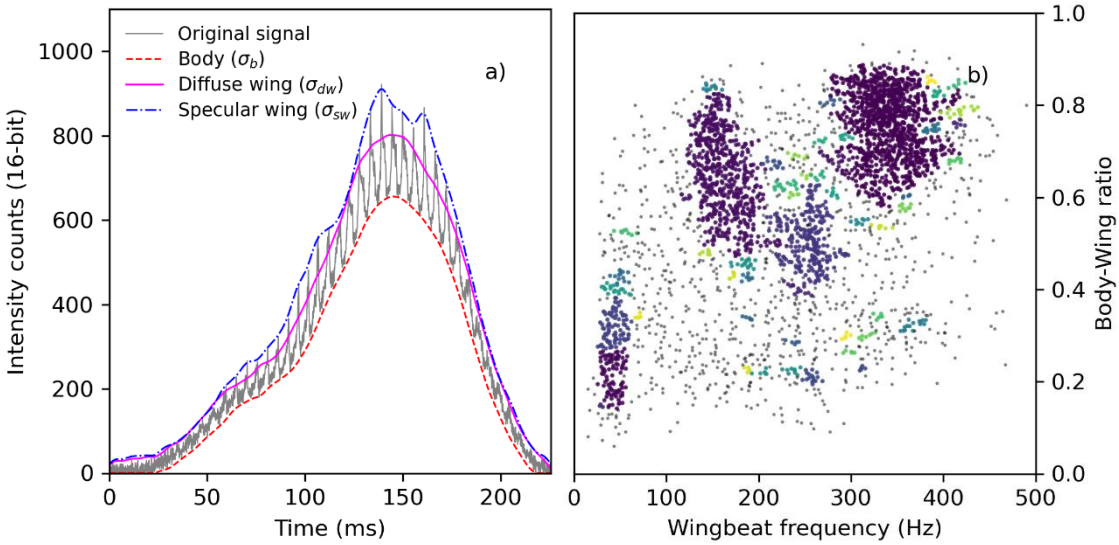


Figure 2. Example of an insect signal and clustering. a) An example of an insect recording from the sensor. The wing beats are visible as modulations on top of the signal. The dashed red, solid magenta and dash-dotted blue curves show the body, diffuse- and specular wing signals respectively. The BWR is the ratio between the magnitude of the body- and diffuse wing signal. b) Clustered insect recordings from a soybean field (Field R) in July. The grey events are too sparse to form clusters and are therefore discarded.

Insects of the same species exhibit similar physical properties, and therefore also similar signal features (Kirkeby, Rydhmer, Cook, & Strand, 2021). Normalization of the feature space is a standard procedure prior to clustering. While BWR values are bound between 0 and 1 by definition (equation 1), WBF values frequencies typically vary between 20 Hz (Jansson et al., 2019) and 1 kHz (Jansson et al., 2019). WBFs were therefore divided by 1000 to produce values falling predominantly between 0 and 1. For clustering, we used the DBSCAN (Density-based spatial clustering of applications with noise) algorithm (Ram et al., 2010) due to its suitability in identifying clusters without a Gaussian distribution assumption (Ester et al., 1996). DBSCAN uses two parameters, the minimum number of insects needed to form a cluster (min_samples) and the merge distance

ϵ, to determine which observations to merge into clusters. Data points too far away from any cluster and too sparsely distributed to form a new cluster are defined as outliers. This method was used to calculate the number of clusters or distinct groups (i.e. richness) and diversity index of cluster groups based on Shannon and Simpson indices.

All insects collected with Malaise traps and sweep nets were classified by order, family, and species when possible. Then species richness (defined as the number of distinct taxonomic species present, independent of abundance), Shannon index, and Simpson index were calculated on the insect samples from both conventional methods for each field in June and July.

The data from the capture methods were randomly divided into two data sets: one used to optimize the DBSCAN clustering algorithm, and the other used for testing. To have a sufficiently large test set, the optimization set was limited to 30% of the data collected. During the optimization, ϵ and min samples were tuned to maximize the Spearman correlation between biodiversity metrics from the sensors and conventional metrics using stochastic gradual descent. This process was repeated for the richness and Shannon and Simpson indices for each of the trapping methods, plus an additional model fitted to the combined species richness from both manual methods. Shannon and Simpson indices were not calculated on the combined dataset since these indices rely on the relative abundance of species, which are not comparable between the two methods.

Optimal parameters could be found that produced significant correlation ($p < 0.05$) for four of the seven comparative measures, however no parameters could be found which satisfactorily modelled the Shannon index from the sweep netting nor the Simpson index for either trapping method.

Spearman-rank correlations between the clustering results calculated from the optical sensor data and the biodiversity measures obtained with the two physical insect field-sampling methods were calculated. Additionally, Analyses of Variance (ANOVA) and TukeyHSD post hoc analyses were conducted to evaluate the impact of sampling month, crop type, and field on richness estimates.

Results

In total, 106,083 insect observations were recorded by the sensors. The Malaise traps collected 14,641 insects, whereas sweep nets collected 15,858 insects. The optical sensors recorded 106 083 insect observations. (Table 1, Figure 3). Moreover, measured insect abundance was uncorrelated between both manual methods and sensors (Figure 4; Malaise trap counts and sweep net counts $r = 0.25$, $p = 0.16$, sensor observations and sweep net counts $r = 0.05$, $p = 0.78$, sensor observations and Malaise trap counts $r = 0.05$, $p = 0.88$).

	SWEEP NET	MALAISE TRAPS	SENSORS
ALFALFA	N=2, μ=3416.5±4472.5	N=2, μ=784.0±1055.0	N=2, μ=4483.0±1479.3
C. COVER CROP	N=8, μ=434.4±232.2	N=10, μ=351.0±242.2	N=9, μ=3408.7±2157.2
CORN	N=6, μ=178.3±185.8	N=4, μ=375.5±339.6	N=3, μ=919.0±235.6
PASTURE	N=3, μ=411.0±274.4	N=3, μ=490.7±595.4	N=3, μ=1676.0±537.8
SORGHUM	N=6, μ=131.3±163.2	N=6, μ=451.0±545.2	N=5, μ=4862.2±2404.9
SOYBEANS	N=10, μ=245.9±245.3	N=12, μ=323.6±385.6	N=12, μ=2861.9±1872.7

***Table 1**. Measured insect abundance per crop and monitoring method. Mean and standard deviation.*

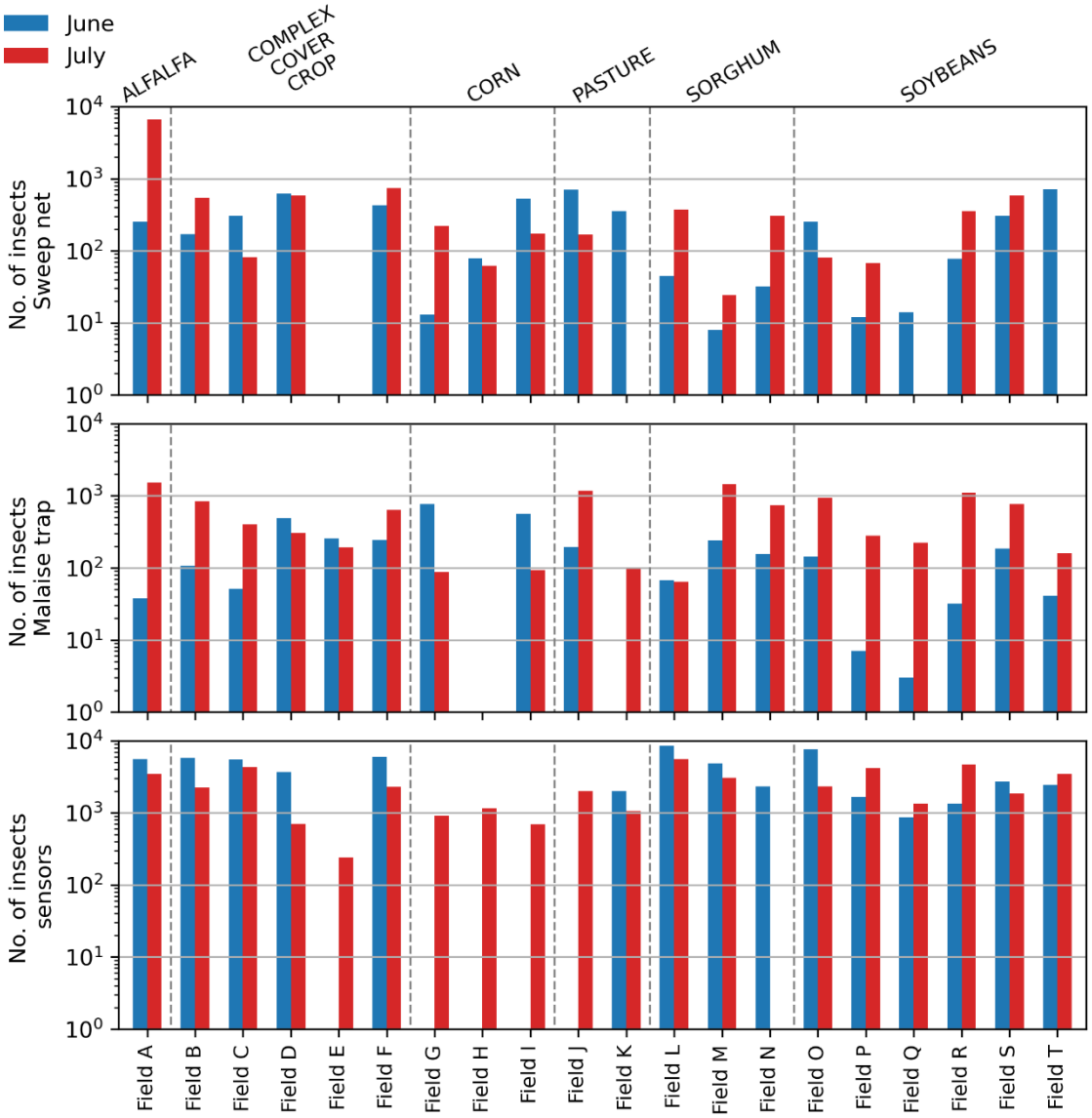


Figure 3. *The number of recorded insect signals (lower) and collected insects in Malaise traps (middle) and sweep nets (lower) per field visit in June and July.*

Comparing the relative insect abundance between orders in manual methods (Figure 5) shows differences in capturing biases. Diptera were most frequently captured in Malaise traps, whereas Hemiptera then Diptera and Psocoptera were more frequently captured with sweep nets. In general, less flight-active insects were more prominent in the sweep net data.

There were no discernible differences in variation between time points from sensors ($F = 0.155$, $Pr = 0.71$) and sweep nets ($F = 0.87$, $Pr = 0.36$). However, Malaise trap abundance showed significantly greater insect densities in July ($\mu = 76.9$, $F = 8.$, $Pr = 0.007$) than June ($\mu = 43.3$). Average abundance per crop is presented in table 1. Crop type was found to impact sweep net abundance ($F = 3.367$, $Pr = 0.01$) but not sensor ($F = 1.76$, $Pr = 0.15$) or Malaise trap ($F = 1.09$, $Pr = 0.44$) abundance estimates. A series of TukeyHSD post hoc analyses found no differences in abundance estimates between sample time points for each field.

For the sensors, the best correlation with the conventional method was found when optimizing the correlation between the number of clusters in the sensors and the combined richness of malaise and sweep nets. The correlation between the sensor estimates and each of the comparative biodiversity metrics are shown in Table 2.

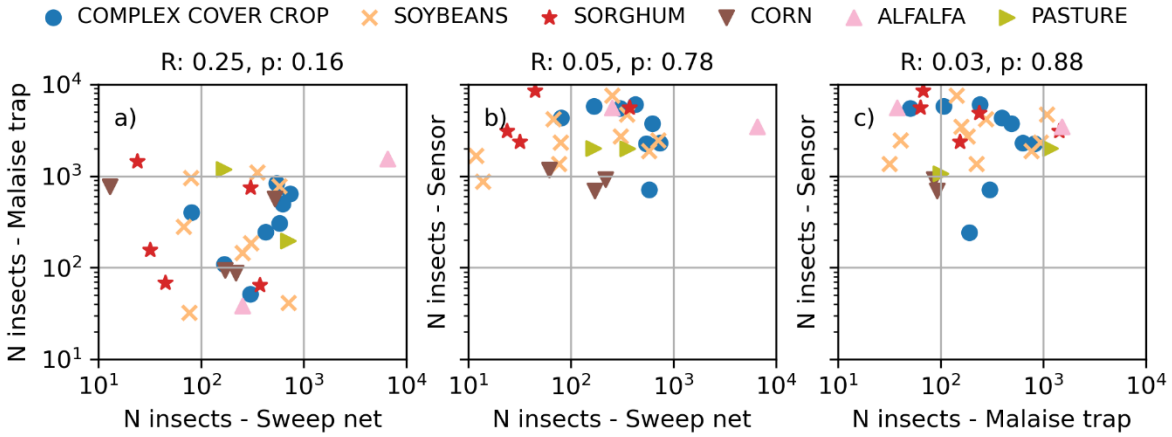


Figure 4. *Scatter plots of measured insect abundance comparing the monitoring methods on a logarithmic scale. a) Scatter plot of the number of insects captured with sweep nets and Malaise traps. b) Scatter plot of the number of insects captured with sweep nets and the number of insect observations recorded by the sensor. c) Scatter plot of the number of insects captured with Malaise traps and the number of insect observations recorded by the sensor. There are no significant correlations between any of the methods.*

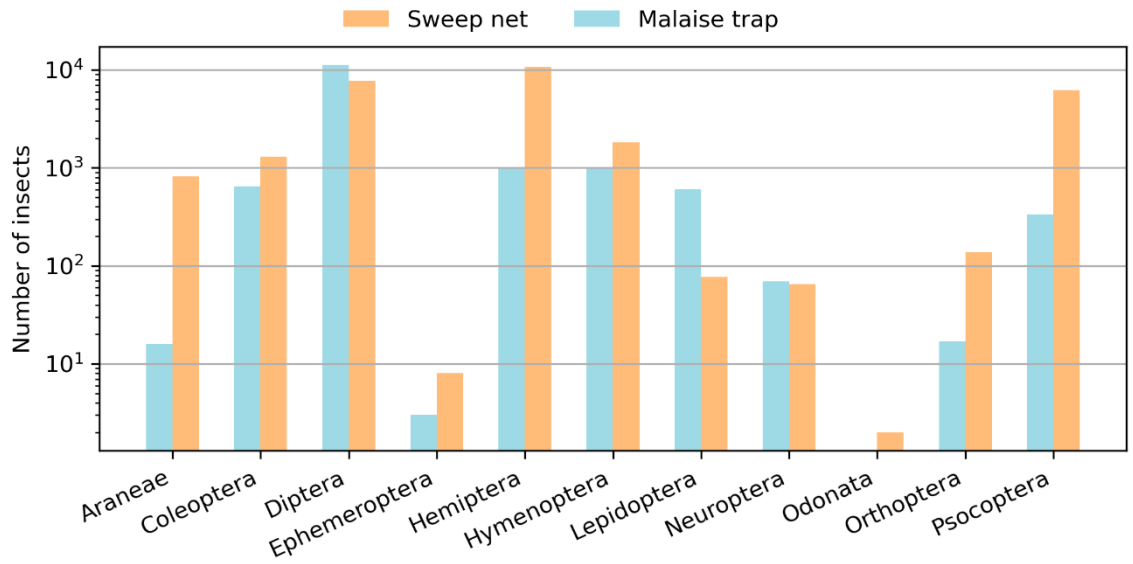


Figure 5. *The number of insects collected with sweep net sampling and Malaise trap monitoring, aggregated by order.*

Per field, the maximum number of clusters was 85 (N=34, μ= 41.1, σ=19.2). The Malaise traps had a maximum richness of 159 species (N = 37, μ = 60.5, σ=39.1) and contained 10 orders, 146 families, and 709 species. The maximum richness observed in the sweep nets was 132 (N=35, μ =47.4, σ=32.7) and contained 11 orders, 149 families, and 664 species. Combined, the collected samples with both field-sampling methods contained 941 different species distributed over 183 insect families and 11 orders.

The three models fitted on sweep net, Malaise and combined species richness are generally comparable (ε: $\mu=8.71\cdot10^{-3}$, $\sigma=8.87\cdot10^{-4}$, min_samples: μ=5.67, σ=0.94). Identical DBSCAN parameters were calculated when the models were fit on the Malaise trap richness and Shannon index. A full list of model parameters is provided as supplementary material (supplementary table 3). As the model fitted on combined richness showed the best correlation with the conventional methods, it was used in the following results.

All species richness metrics were correlated (Figure 6). The weakest correlation was between Malaise trap and sweep net richness metrics (R = 0.36, p = 0.046). The correlation between the number of clusters found in the sensor data and the conventional models was strongest for the combined richness, which was what the model

was fitted on (R= 0.55, p = 0.012; Figure 6d). Significant yet weaker correlations were also found between the model and the Malaise trap and sweep net richness respectively (R = 0.52, p = 0.014; R = 0.48, p = 0.028).

Fitting data	Correlation coefficients (R, p-value)						
	Richness			Shannon		Simpson	
	Malaise trap	Sweep net	Combined	Malaise trap	Sweep net	Malaise trap	Sweep net
Richness Malaise	**0.42, 0.050**	0.26, 0.246	0.37, 0.112	**0.44, 0.044**	-0.22, 0.359	0.35, 0.114	0.01, 0.964
Richness sweep net	**0.45, 0.035**	**0.44, 0.046**	**0.47, 0.035**	0.11, 0.612	0.20, 0.378	0.09, 0.702	0.16, 0.475
Richness combined	**0.52, 0.014**	**0.48, 0.028**	**0.55, 0.012**	0.21, 0.344	0.20, 0.394	0.14, 0.534	0.14, 0.537
Shannon Malaise	**0.42, 0.05**0	0.26, 0.246	0.37, 0.112	**0.44, 0.044**	-0.22, 0.359	0.35, 0.114	0.01, 0.964
Shannon sweep net	**0.45, 0.037**	**0.37, 0.097**	**0.46, 0.040**	0.27, 0.221	0.22, 0.339	0.17, 0.459	0.08, 0.733
Simpson Malaise	**0.44, 0.041**	0.32, 0.159	0.43, 0.061	-0.16, 0.484	0.03, 0.893	0.02, 0.913	0.06, 0.797
Simpson sweep net	0.05, 0.837	-0.13, 0.569	-0.07, 0.774	**-0.73, 0.005**	0.07, 0.817	-0.04, 0.846	0.12, 0.598

Table 2. *Correlations between the automated sensors biodiversity metrics and those obtained from Malaise trap and sweep net collections on the test set. Rows in the table denote which data was used to fit the clustering algorithm, whereas columns indicate which parameters the obtained correlations refer to. Significant correlations with a p-value below 0.05 are marked in bold.*

No correlation was found when comparing sensor richness to any ecosystem services (Table 3). Manual sampling methods were typically not correlated with ecosystem services with one exception. Sweep net species richness was correlated with the percent of waxworms predated.

Richness metrics	% Waxworms Predated	Total # of predators	Johnsongrass predation	Pigweed predation	Lambsquarter predation	All seed predation
Sweep net	0.49, 0.04	0.34, 0.17	-0.37, 0.13	0.05, 0.85	-0.04, 0.89	-0.07, 0.78
Malaise trap	-0.16, 0.52	-0.13, 0.61	-0.16, 0.52	-0.36, 0.15	-0.40, 0.10	-0.45, 0.06
Total - richness	0.21, 0.43	0.16, 0.56	-0.39, 0.14	-0.26, 0.33	-0.32, 0.23	-0.41, 0.11
Sensor clusters	-0.11, 0.66	-0.26, 0.29	0.14, 0.56	-0.11, 0.66	-0.15, 0.54	-0.11, 0.66

Table 3. *Correlation table between richness metrics calculated from sweep nets, Malaise traps, combined conventional methods, and sensor clusters (novel biodiversity metric) compared to ecosystem services of percent waxworm predation, total number of predators, Johnsongrass predation, Pigweed predation, Lambsquarter predation, and all seed predation.*

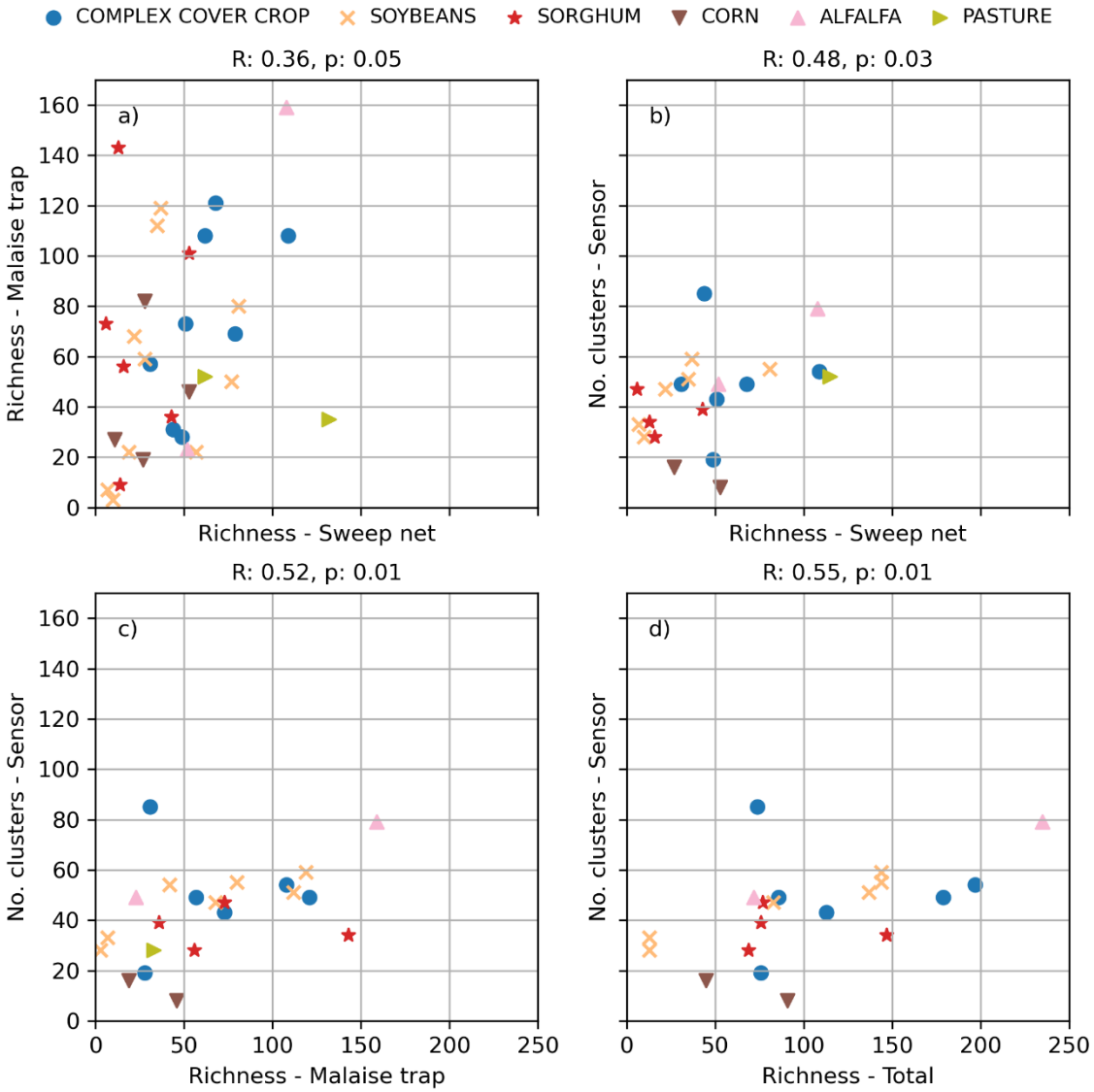


Figure 6. *Scatter plots and Spearman correlations for the richness estimations across all models. The sensor results are from the model fitted to the total richness in both Malaise traps and sweep nets. a) Malaise traps vs. sweep net samples, b) Sensors vs. sweep net samples, c) Sensors vs. Malaise trap samples, and d) Sensors vs. total richness across traps and sweeps.*

Discussion

The sensor recorded a greater total number of insects observations, almost one order of magnitude, than both Malaise traps and sweep nets. This difference in measurement rate may be explained as a combination of two factors: measurement period and measurement methodology. Firstly, the sensors continuously monitored each field for three times longer than the Malaise traps, the other continuous monitoring method. Since sweep nets provide point-in-time measurements, the methods are not directly comparable. Secondly, the sensor does not rely on trapping to record an insect signal. This means both that individual insects may be measured more than once as they fly in front of the optical sensor. Both factors may result in a higher insect measurement rate of sensors compared to conventional methodologies. This tendency has been observed in previous work, for example a similar sensor reported 19 times as many insects as water traps collecting over the same period (Rydhmer et al., 2021), and optical sensors with greater measurement volumes report recording tens of thousands of insect flights per day (Brydegaard et al., 2020).

One challenge with the sensor's dataset was the high proportion of noise signals, primarily thought to be caused by plant interference. Of the total 1,057,115 signals recorded by the sensor, only ca 10% were classified as insect signals and included in the analysis. We therefore expect that the reported number of insect observations is artificially increased. The signals generated by insects and plants moving through the sensor's measurement volume are very different. While we believe the proprietary neural network model used to filter observations is highly accurate, even a small percentage of misclassifications of large number of noise events will inevitably inflate the total count. Most of these misclassified events have do not show any modulations in time, (since plant interference has no high frequency components) and are therefore assumed to be removed by filtering the data on frequency, and body-wing-ratio. Despite these efforts, it is likely that some noise remains. However, we also believe that the strong signals generated by vegetation will obscure weaker signals generated by small insects. This would in contrast reduce the total insect count.

Our results suggest that the sensor-derived metric is correlated with conventional estimates of biodiversity. This indicates that metrics derived from optical sensors have the potential to provide accurate and autonomous measurements of insect species richness. Still, future work is needed to evaluate the extent to which the metric may be generalized across agroecosystems outside our study area and to other terrestrial ecosystems. While a species metric that does not characterize the composition of the insect community presents some added difficulty to stakeholders in formulating a targeted response, significant and growing evidence suggests that biodiversity itself is correlated with greater ecosystem functions such as pest control (J. G. Lundgren & Fausti, 2015).

The lack of correlation of abundance and the three methods (Figure 4) is surprising as previous work has shown correlations between sensor measurements and water traps for insect abundance (Rydhmer et al., 2021). However, while sweep netting occurred in conjunction with setting up or taking down the Malaise traps, these efforts were substantially less correlated with the setup of the optical sensors: to the nearest 3 days in June and nearest 22 days in July. The lack of abundance correlation between the Malaise traps and the sensors may be due to the long period between the monitoring sessions at each site. Insect flight activity is heavily influenced by the weather, or the seasonal differences between the beginning and end of July – both of which may also explain the significance of month on Malaise trap data. An additional factor may be the previously mentioned high noise composition of the recorded signals due to plant interference. During cleaning of this dataset, it is possible that variations in the relative degree of noise signals between fields (e.g. because of different crop heights and stiffness) has resulted in more data loss from noisier fields, thus introducing a systematic error in abundance measurements for the sensor data.

Due to the similarities between the sensors and Malaise traps (both monitoring flying insects over extended periods of time), we expected stronger correlations between the sensors and Malaise traps compared to the sensors and sweep nets. However, those results were comparable. The results were less clear for the correlations between species diversity indices (Shannon and Simpson). The models fitted on malaise trap data yielded identical models for the richness and Shannon index. This is most likely due to the co-correlation between the richness and Shannon index in the malaise trap ($R=0.6$, $p=0.01$, Supplementary table 2). Other curiosities, such as the negative correlation with the Malaise trap Shannon index achieved when fitting on the sweep net Simpson index are also assumed to be the results of co-correlations between the conventional methods. A full table of all co-correlations is included as supplementary material.

An entomological radar group called BioDAR is aiming to use libraries of insect radar signals for functional group classifications for high flying migratory insects at a regional scale (Lukach et al., 2022). Similar estimates

of insect functional groups might be similarly inferred from optical sensor recordings for all flying insects on a field scale. Similarly, vertical looking radar is used to classify insects into higher level taxonomic groups such as Order or even Genus (Chapman et al., 2002; Stefanescu et al., 2013; Wood et al., 2009). It seems likely that similar or even higher precision can be achieved by further analysis of optically recorded data. Future work could focus on identifying these groups, determining functional biodiversity, and quantifying their ecosystem services.

The current study shows a single instance of correlation between richness and a measure of ecosystem services. Greater species richness does not always translate into an increase in functional biodiversity or ecosystem services, as there is often ecological redundancy (Greenop et al., 2018). The lack of a relationship may also reflect different ecological interactions among species in the upper canopy versus above canopy level. These questions can be further explored in future work when the sensors ability to estimate functional biodiversity has been developed.

Acknowledgments

We thank the USDA Cheney Lake Conservation District (Lisa French), Understanding Ag (Ray Archuleta) for logistical support and field assistance. Tom Rabaey (General Mills) provided helpful feedback and advice to guide this project. We thank Kevin James Knagg and Mads Fogtmann from FaunaPhotonics A/S for facilitating this work. We thank Ecdysis Foundation field and lab technicians for collecting, processing, and Dr. Kelton D. Welch for identifying invertebrate specimens. We thank the farmers cooperating in the General Mills regenerative Agriculture Programs who granted us access and permission to sample their farm fields.

References

Bick, E., Sigsgaard, L., Torrance, M. T., Helmreich, S., Still, L., Beck, B., Rashid, R. El, Lemmich, J., Nikolajsen, T., & Cook, S. M. (2023). Dynamics of pollen beetle (Brassicogethes aeneus) immigration and colonisation of oilseed rape (Brassica napus) in Europe. *Pest Management Science*. https://doi.org/10.1002/PS.7538

Brydegaard, M., Jansson, S., Malmqvist, E., Mlacha, Y. P., Gebru, A., Okumu, F., Killeen, G. F., & Kirkeby, C. (2020). Lidar reveals activity anomaly of malaria vectors during pan-African eclipse. *Science Advances*, *6*(20). https://doi.org/10.1126/sciadv.aay5487

Chapman, J. W., Reynolds, D. R., Smith, A. D., Riley, J. R., Pedgley, D. E., & Woiwod, I. P. (2002). High-altitude migration of the diamondback moth Plutella xylostella to the U.K.: a study using radar, aerial netting, and ground trapping. *Ecological Entomology*, *27*(6), 641–650. https://doi.org/10.1046/j.1365-2311.2002.00472.x

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters... - Google Scholar. *Kdd*, *96*(34), 226–231. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C50&q=A+Density-Based+Algorithm+for+Discovering+Clusters+in+Large+Spatial+Databases+with+Nois&btnG=#d=gs_cit&t=1677591470201&u=%2Fscholar%3Fq%3Dinfo%3A-KybkyxcGYIJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den

Garcia, K., Olimpi, E. M., M'Gonigle, L., Karp, D. S., Wilson-Rankin, E. E., Kremen, C., & Gonthier, D. J. (2023a). Semi-natural habitats on organic strawberry farms and in surrounding landscapes promote bird biodiversity and pest control potential. *Agriculture, Ecosystems & Environment*, *347*, 108353. https://doi.org/10.1016/J.AGEE.2023.108353

Garcia, K., Olimpi, E. M., M'Gonigle, L., Karp, D. S., Wilson-Rankin, E. E., Kremen, C., & Gonthier, D. J. (2023b). Semi-natural habitats on organic strawberry farms and in surrounding landscapes promote bird biodiversity and pest control potential. *Agriculture, Ecosystems & Environment*, *347*, 108353.

Gardner, T. A., Barlow, J., Araujo, I. S., Ávila-Pires, T. C., Bonaldo, A. B., Costa, J. E., Esposito, M. C., Ferreira, L. V, Hawes, J., & Hernandez, M. I. M. (2008). The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters*, *11*(2), 139–150.

Gebru, A., Jansson, S., Ignell, R., Kirkeby, C., Prangsma, J. C., & Brydegaard, M. (2018). Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight. *Journal of Biophotonics*, *11*(8). https://doi.org/10.1002/jbio.201800014

Geiger, M. F., Moriniere, J., Hausmann, A., Haszprunar, G., Wägele, W., Hebert, P. D. N., & Rulik, B. (2016a). Testing the Global Malaise Trap Program – How well does the current barcode reference library identify flying insects in Germany? *Biodiversity Data Journal*, *4*(4). https://doi.org/10.3897/BDJ.4.E10671

Geiger, M. F., Moriniere, J., Hausmann, A., Haszprunar, G., Wägele, W., Hebert, P. D. N., & Rulik, B. (2016b). Testing the Global Malaise Trap Program–How well does the current barcode reference library identify flying insects in Germany? *Biodiversity Data Journal*, *4*.

Genoud, A. P., Gao, Y., Williams, G. M., & Thomas, B. P. (2019). Identification of gravid mosquitoes from changes in spectral and polarimetric backscatter cross sections. *Journal of Biophotonics*, *12*(10), e201900123.

Greenop, A., Woodcock, B. A., Wilby, A., Cook, S. M., & Pywell, R. F. (2018). Functional diversity positively affects prey suppression by invertebrate predators: a meta-analysis. *Wiley Online Library*, *99*(8), 1771–1782. https://doi.org/10.1002/ecy.2378

Jansson, S., Gebru, A., Ignell, R., & Abbott, j. (2019). Correlation of mosquito wing-beat harmonics to aid in species classification and flight heading assessment. *European Conference on Biomedical Optic*. https://opg.optica.org/abstract.cfm?uri=ecbo-2019-11075_24

Kirkeby, C., Rydhmer, K., Cook, S. M., Strand, A., Torrance, M. T., Swain, J. L., Prangsma, J., Johnen, A., Jensen, M., Brydegaard, M., & Græsbøll, K. (2021). Advances in automatic identification of flying insects

using optical sensors and machine learning. *Scientific Reports*, *11*(1), 1555. https://doi.org/10.1038/s41598-021-81005-0
Kirkeby, C., Rydhmer, K., Cook, S., & Strand, A. (2021). Advances in automatic identification of flying insects using optical sensors and machine learning. *Scientific Reports*, *11*(1), 1555. https://www.nature.com/articles/s41598-021-81005-0
Krishna Krishnamurthy, P., & Francis, R. A. (2012). A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation 2012 21:8*, *21*(8), 1901–1919. https://doi.org/10.1007/S10531-012-0306-2
LaCanne, C. E., & Lundgren, J. G. (2018a). Regenerative agriculture: Merging farming and natural resource conservation profitably. *PeerJ*, *2018*(2), e4428. https://doi.org/10.7717/PEERJ.4428/SUPP-1
LaCanne, C. E., & Lundgren, J. G. (2018b). Regenerative agriculture: merging farming and natural resource conservation profitably. *PeerJ*, *6*, e4428.
Landis, D. A., Gardiner, M. M., Van Der Werf, W., & Swinton, S. M. (2008). Increasing corn for biofuel production reduces biocontrol services in agricultural landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20552–20557. https://doi.org/10.1073/PNAS.0804951106/SUPPL_FILE/0804951106SI.PDF
Lukach, M., Dally, T., Evans, W., Hassall, C., Duncan, E. J., Bennett, L., Addison, F. I., Kunin, W. E., Chapman, J. W., & Neely, R. R. (2022). The development of an unsupervised hierarchical clustering analysis of dual-polarization weather surveillance radar observations to assess nocturnal insect abundance and diversity. *Remote Sensing in Ecology and Conservation*, *8*(5), 698–716. https://doi.org/10.1002/RSE2.270
Lundgren, J. G., & Fausti, S. W. (2015). Trading biodiversity for pest problems. *Science Advances*, *1*(6), e1500558.
Lundgren, J. G., Shaw, J. T., Zaborski, E. R., & Eastman, C. E. (2006). The influence of organic transition systems on beneficial ground-dwelling arthropods and predation of insects and weed seeds. *Renewable Agriculture and Food Systems*, *21*(4), 227–237.
Lundgren, J., Shaw, J., Zaborski, E., & Eastman, C. (2006). The influence of organic transition systems on beneficial ground-dwelling arthropods and predation of insects and weed seeds. *Renewable Agriculture and Food Systems*, *21*(4), 227–237. https://pubag.nal.usda.gov/catalog/49820
Montgomery, G. A., Belitz, M. W., Guralnick, R. P., & Tingley, M. W. (2021). Standards and Best Practices for Monitoring and Benchmarking Insects. *Frontiers in Ecology and Evolution*, *8*, 579193. https://doi.org/10.3389/FEVO.2020.579193/BIBTEX
Morris, R. F. (1960). Sampling insect populations. *Annual Review of Entomology*, *5*(1), 243–264.
Preti, M., Verheggen, F., & Angeli, S. (2021). Insect pest monitoring with camera-equipped traps: strengths and limitations. *Journal of Pest Science*, *94*(2), 203–217. https://doi.org/10.1007/S10340-020-01309-4/TABLES/2
Ram, A., Jalal, S., & Kumar, m. (2010). A density based algorithm for discovering density varied clusters in large spatial databases. *International Journal of Computer Applications*, *3*(6), 1–4. https://doi.org/10.13140/RG.2.1.4420.1448
Rydhmer, K., Bick, E., Still, L., Strand, A., Luciano, R., Helmreich, S., Beck, B. D., Grønne, C., Malmros, L., & Poulsen, K. (2022). Automating insect monitoring using unsupervised near-infrared sensors. *Scientific Reports*, *12*(1), 1–11.
Rydhmer, K., Bick, E., Still, L., Strand, A., Luciano, R., Helmreich, S., Beck, B., Grønne, C., Malmros, L., Poulsen, K., Elbæk, F., Brydegaard, M., Lemmich, J., & Nikolajsen, T. (2021). *Automating insect monitoring using unsupervised near-infrared sensors*. https://arxiv.org/abs/2108.05435v1
Sánchez-Bayo, F., & Wyckhuys, K. A. G. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation*, *232*, 8–27. https://doi.org/10.1016/J.BIOCON.2019.01.020
Shortall, C. R., Moore, A., Smith, E., Hall, M. J., Woiwod, I. P., & Harrington, R. (2009). Long-term changes in the abundance of flying insects. *Insect Conservation and Diversity*, *2*(4), 251–260.
Silva, D. F., De Souza, V. M. A., Batista, G. E., Keogh, E., & Ellis, D. P. W. (2013). Applying machine learning and audio analysis techniques to insect recognition in intelligent traps. *2013 12th International Conference on Machine Learning and Applications*, *1*, 99–104.
Stefanescu, C., Páramo, F., Åkesson, S., Alarcón, M., Ávila, A., Brereton, T., Carnicer, J., Cassar, L. F., Fox, R., Heliölä, J., Hill, J. K., Hirneisen, N., Kjellén, N., Kühn, E., Kuussaari, M., Leskinen, M., Liechti, F., Musche, M., Regan, E. C., … Vlinderstichting, D. (2013). Multi-generational long-distance migration of

insects: studying the painted lady butterfly in the Western Palaearctic. *Wiley Online Library*, *140*(4), 474–486. https://doi.org/10.1111/j.1600-0587.2012.07738.x

Tilman, D., May, R. M., Lehman, C. L., & Nowak, M. A. (1994). Habitat destruction and the extinction debt. *Nature*, *371*(6492), 65–66.

Wägele, J. W., Bodesheim, P., Bourlat, S. J., Denzler, J., Diepenbroek, M., Fonseca, V., Frommolt, K.-H., Geiger, M. F., Gemeinholzer, B., & Glöckner, F. O. (2022). Towards a multisensor station for automated biodiversity monitoring. *Basic and Applied Ecology*, *59*, 105–138.

Wood, C. R., Reynolds, D. R., Wells, P. M., Barlow, J. F., Woiwod, I. P., & Chapman, J. W. (2009). Flight periodicity and the vertical distribution of high-altitude moth migration over southern Britain. *Bulletin of Entomological Research*, *99*, 525–535. https://doi.org/10.1017/S0007485308006548

Yang, L. H., & Gratton, C. (2014). Insects as drivers of ecosystem processes. *Current Opinion in Insect Science*, *2*, 26–32.

511 **Supplementary material**

512

		Sensor		Malaise trap		Sweep net	
Field name	Crop type	June	July	June	July	June	July
A	ALFALFA	5529	3437	38	1530	254	6579
B	CORN	-	1158	-	-	79	62
C	CORN	5739	2244	108	828	169	546
D	CORN	-	687	561	93	525	172
E	CORN	-	912	761	87	13	219
F	HAYLAND	3718	704	493	303	626	583
G	ANNUAL PASTURE	1657	4162	7	278	12	68
H	ANNUAL PASTURE	5454	4290	51	398	303	81
I	ANNUAL PASTURE	6008	2281	243	636	427	740
J	SORGHUM	4845	3080	239	1437	8	24
K	SORGHUM	8492	5545	68	64	45	375
L	SORGHUM	2349	-	155	743	32	304
M	SORGHUM	7524	2310	144	948	254	80
N	PERENNIAL PASTURE	-	1993	195	1176	708	167
O	PERENNIAL PASTURE	1980	1055	-	101	358	-
P	SOYBEAN	865	1349	3	225	14	-
Q	SOYBEAN	1349	4679	32	1090	77	354
R	SOYBEAN	2443	3442	41	160	713	-
S	SOYBEAN	2696	1867	184	771	307	580
T	SUDANGRASS	-	240	258	192	-	-

513 **Supplementary Table 1.** *Crop type and insects abundance observed in each field in June and July across all*
514 *three methods.*
515

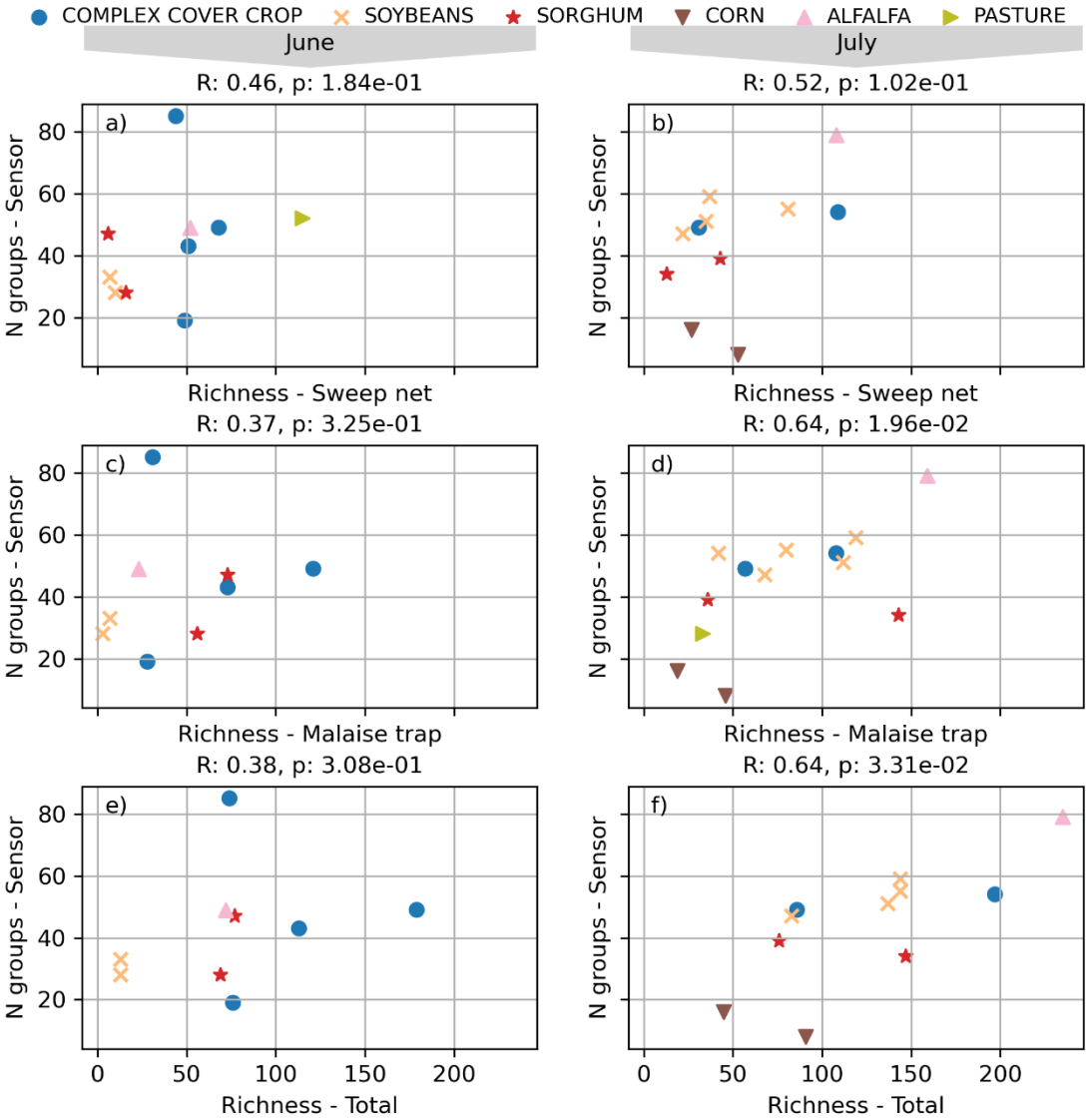
516

	Simpson Sensor	Shannon Sensor	No. clusters sensor	Simpson Malaise	Shannon Malaise	Richness Malaise	Simpson sweepnet	Shannon sweepnet	Richness sweep net	Richness combined
Simpson Sensor		0.95, 0.0	0.62, 0.0	0.14, 0.53	0.28, 0.2	0.42, 0.05	0.14, 0.54	0.18, 0.44	0.34, 0.13	0.42, 0.07
Shannon Sensor	0.95, 0.0		0.67, 0.0	0.1, 0.65	0.21, 0.34	0.41, 0.06	0.2, 0.39	0.2, 0.39	0.3, 0.19	0.41, 0.07
No. clusters sensor	0.62, 0.0	0.67, 0.0		-0.05, 0.82	0.26, 0.25	0.52, 0.01	0.2, 0.4	0.36, 0.11	0.48, 0.03	0.55, 0.01
Simpson Malaise	0.14, 0.53	0.1, 0.65	-0.05, 0.82		0.61, 0.0	-0.03, 0.87	-0.17, 0.34	-0.31, 0.09	-0.21, 0.24	-0.23, 0.22
Shannon Malaise	0.28, 0.2	0.21, 0.34	0.26, 0.25	0.61, 0.0		0.59, 0.0	-0.33, 0.07	-0.14, 0.45	0.16, 0.39	0.4, 0.02
Richness Malaise	0.42, 0.05	0.41, 0.06	0.52, 0.01	-0.03, 0.87	0.59, 0.0		-0.14, 0.45	0.08, 0.68	0.36, 0.05	0.85, 0.0
Simpson sweepnet	0.14, 0.54	0.2, 0.39	0.2, 0.4	-0.17, 0.34	-0.33, 0.07	-0.14, 0.45		0.72, 0.0	0.02, 0.89	0.36, 0.68
Shannon sweepnet	0.18, 0.44	0.2, 0.39	0.36, 0.11	-0.31, 0.09	-0.14, 0.45	0.08, 0.68	0.72, 0.0		0.58, 0.0	0.35, 0.05
Richness sweep net	0.34, 0.13	0.3, 0.19	0.48, 0.03	-0.21, 0.24	0.16, 0.39	0.36, 0.05	0.02, 0.89	0.58, 0.0		0.73, 0.0
Richness combined	0.42, 0.07	0.41, 0.07	0.55, 0.01	-0.23, 0.22	0.4, 0.02	0.85, 0.0	-0.06, 0.74	0.35, 0.05	0.73, 0.0	

517
518 **Supplementary Table 2.** *Co-correlations of all biodiversity metrics*.
519
520

Fitting data	Richness Malaise	Richness sweep net	Richness combined	Shannon Malaise	Shannon sweep net	Simpson Malaise	Simpson sweep net
ε	0.007467	0.009467	0.0092	0.007467	0.0104	0.0084	5.33E-04
min_ samples	7	5	5	7	7	3	3

Supplementary Table 3: *Optimal model parameters for each fitted biodiversity metric.*



Supplementary Figure 2. A scatterplot depicting the correlation of the species richness metrics at each field, separated by the June and July timepoints.

Paper V: Photonic sensors for comparative insect abundance and diversity in distinct habitats

Photonic sensors for comparative insect abundance and diversity in distinct habitats

Klas Rydhmer[1,2], Samuel Jansson[1], Laurence Still[1], Brittany D. Beck[1], Vasileia Chatzaki[1], Karen Olsen[1], Bennett Van Hoff[1], Christoffer Grønne[1], Jakob Klinge Meier[1], Marta Montaro[1], Inger Kappel Schmidt[2], Carsten Kirkeby[1,3], Henrik G. Smith[4], Mikkel Brydegaard[1,4,5]

1 FaunaPhotonics, Støberigade 14, 2450 Copenhagen, Denmark

2 Department of Geosciences and Natural Resource Management, Faculty of Science, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark

3 Department of Veterinary and Animal Sciences, Section for Production, Nutrition and Health, University of Copenhagen, 1870 Frederiksberg C, Denmark

4 Department of Biology & Centre of Environmental and Climate Science, Lund University, Sölvegatan 37, 223 62 Lund, Sweden

5 Department of Physics, Lund University, Sölvegatan 14c. 22362 Lund, Sweden

Abstract

To mitigate the ongoing declines in insect biodiversity, there is a need of efficient yet accurate monitoring methods. The use of traditional catch-based survey methods is constrained by the costs and the need for expertise for manual taxonomic identification. Emerging methods, such as eDNA and robotic sorting, have the potential to reduce workload, but still requires resource-intensive sample collection in the field. Recently, remote sensing methods such as photonic sensors have shown promise for recording large numbers of insect observations. However, accurately determining species composition in the collected data remains challenging.

In this study, we investigate the potential of photonic sensors for quantifying species richness in the field and compare the results with estimates based on conventional Malaise traps at five sites. Firstly, we evaluated two unsupervised clustering methods using a library of measured insect signals from known species. Then we correlated the estimated number of clusters with the species richness assessment by the Malaise trapping. Our results demonstrate that both clustering methods perform well when compared to the Malaise traps, indicating the potential of automated insect biodiversity monitoring. This offers the possibility of more efficient but still accurate methods for studying insect biodiversity with broader temporal and spatial coverage.

Keywords: Insects, Biodiversity, Clustering, Photonics, Entomology, Ecology, Modulation Spectroscopy

1. Introduction/Background

Declines of terrestrial insects in response to anthropogenic drivers [1]–[4] [5], [6] calls for documentation of trends and drivers. Rapid assessment of insect communities is needed to document such declines of insect biomass and diversity, understand its drivers, and evaluate strategies to mitigate them.

Safeguarding insect abundance and diversity is important for many reasons. Negative population trends may ultimately result in the loss of species, violating the conservation promises in the Convention of Biological Diversity [7]. Insects are entangled in trophic webs and vital for ecosystem functions[8]. Additionally, insects contribute with ecosystem services important to humans, such as pollination, biological pests control, decomposition, and recreational values [9], [10]. These functions and services depend on specific species and thus their abundance, but are also often enhanced or stabilized by higher diversity [10], [11]. However, insects also play other roles, by acting as pests [12] or disease vectors of human diseases [13].

Conventional catch-based survey methods, such as sweep netting or Malaise traps, are time consuming and resource intensive, both during data collection and the following taxonomic identification [13], [14]. The taxonomical identification to family or species level has historically been dependent on microscopic morphological taxonomic identification by experts. This laborious process hampers rapid evaluation of mitigation measures, and the sampling time resolution is insufficient to capture species weather preferences. Promising methods are in development including robotic sorting [15], [16] and meta-barcoding [17], [18], although these methods still require time-consuming collection of specimens. A challenge for eDNA methods is also to provide abundance estimates [19], although quantitative techniques are emerging [20]. In protected areas where there is an increasing need for monitoring, the destructive nature of catch based methods might also make them difficult to implement [21].

Automated monitoring approaches can provide data with high temporal resolution at low cost [4], [22]. Emerging technologies include e-traps [23], [24], as well as non-destructive methods such as acoustic [25], [26], machine vision [16], [27], [28] and photonic sensors [29]–[32]. Photonic monitoring of insects has been demonstrated over seasons [32], [33] and through wingbeat modulation characteristics, dozens of insect groups can be differentiated *in situ* [34], [35]. Some of these methods could allow efficient retrieval of data with high temporal and spatial resolution across families, but it is important that their ability to accurately reflect true variation in insect abundance and diversity is evaluated.

In this work, we aimed to demonstrate the utility of photonic sensors to generate proxies of biodiversity. We used commercial photonic sensors that emit modulated infrared light and records the backscatter from insects flying through the measurement volume. Using two different clustering algorithms, we estimated the number of unique signals. We developed models using data from from flight cages with known species and evaluated their performance by comparing online *in situ* field estimates at five different sites with concurrent results derived from catches by conventional Malaise traps. The structure of the study is outlined in Fig. 1.

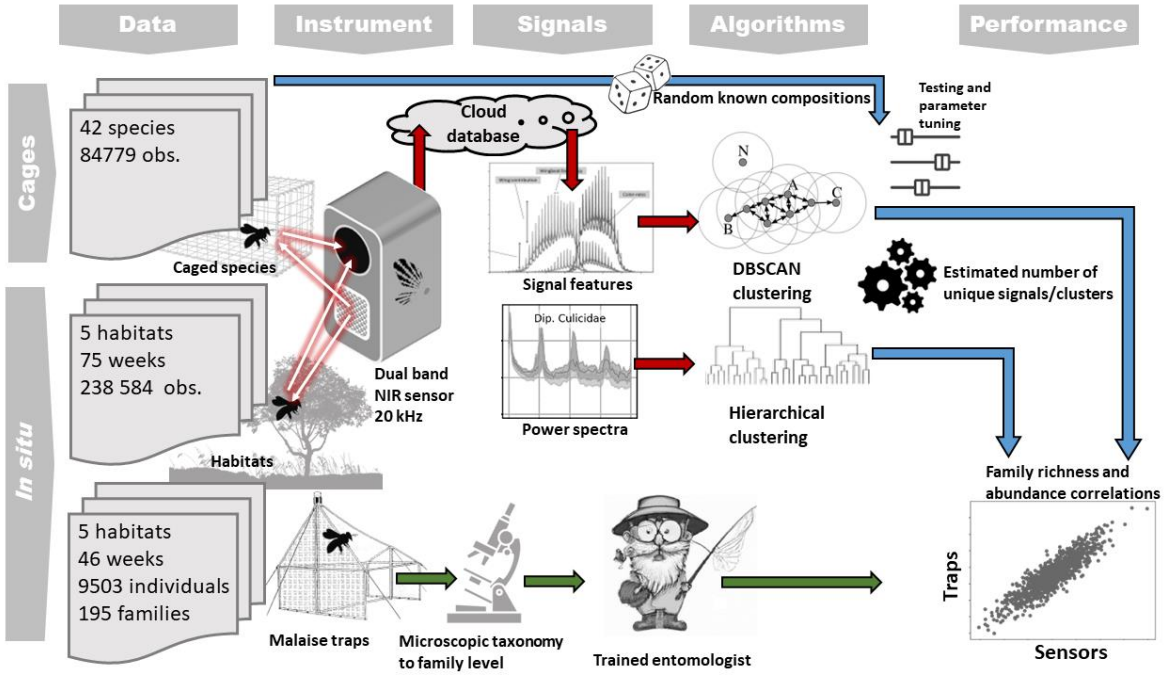


Figure 1. *Experiment flowchart. The algorithms were tuned on a dataset of signals from known species in flight cages. The best clustering algorithm parameters were then evaluated on data collected from sensors in the field, alongside Malaise traps. The number of clusters found by the algorithms in the sensor data was compared to the number of taxonomic families in the Malaise traps.*

2. Experiment design and material

In this study, we evaluated two clustering methods of photonic sensor data for their potential use in biodiversity monitoring. To tune and validate the two clustering methods, we used a library of insect signals from known species recorded in flight cages. Subsequently, the tuned clustering methods were applied to data recorded at multiple field sites. In the same sites, we concurrently captured insects using Malaise traps that were emptied weekly. The richness estimates from photonic data, i.e. the number of identified clusters, were finally evaluated against the family richness (number of insect families present in the sample) estimated through morphological analysis of catches from Malaise traps.

2.1. Insect sensor

The study is based on a recently developed commercial insect sensor (Volito, FaunaPhotonics, Denmark), previously described by Rydhmer et al. [29]. The sensor emits modulated infrared light at 810 and 970 nm and records the backscatter from insects flying through the measurement volume with a quadrant photodetector. The instrument has a sampling bandwidth of 5 kHz and the measurement volume extends around 1 m from the sensors and comprise up to 70 l of air. Recorded insect observations are automatically extracted by the sensor and transmitted to a cloud platform via 4G mobile network for further analysis.

In the cloud platform, a proprietary neural network is used to classify each recording as either an insect observation or noise. The classification algorithm is applied to each quadrant and band individually as described more extensively in the supplementary material.

2.2. Collection of signals from single species cages

Labelled reference data was recorded in flight cages with an approximate volume of 1 m^3. During recordings, up to 100 insects from one species at a time, depending on flight activity, were inserted in the cages and recorded over several days. Water and species appropriate nutrients were present in the cages. In this work, we included 84779 individual insect observations from 42 different species, representing 27 families.

We selected reference species for the flight cages with the aim of representing a wide range of morphological and taxonomic diversity, but selection was constrained by commercial and seasonal availability. Insects were collected in the field, provided by academical partners or acquired from commercial sources. The number of unique observations from each species varied from 653 to 5021 with an average of 2018 observations per species. A table of the included species is provided as supplementary material.

2.3. Field data collection

Sensors were installed within 5 m of conventional Malaise traps (BT1001, Megaview Science co. LTD, Taiwan) at five sites in Denmark and southern Sweden (Table 2). The habitats included an oilseed rape field [29], grazed pasture grasslands [36], protected spruce forest [37], a moist meadow surrounded by deciduous forest [38]. The measurements spanned from April 2020 to November 2020, however, with multiple interruptions resulting from technical or logistical challenges. Malaise traps were emptied weekly and collected insects were stored in ethanol until identification. An overview of the field sites is presented in table 3. During identification, the catch was randomly subsampled by a factor 6 and identified to family level by the authors using microscope and taxonomic keys.

We collected 78 Malaise trap samples from the field sites, but limited the number of samples identified to match available resources. We initially selected 50 samples covering a wide range of abundances. However, 10 samples were later discarded due to technical problems with the sensors, leaving 40 samples available for time-matched comparisons between sensors and traps.

3. Photonic signal processing

3.1 Signal features

Insect signals were recorded as dual-band time series sampled at 20 kHz (5 kHz bandwidth). An example of a recorded insect observation from one detector quadrant is presented in Figure 2. From these observations, we isolated the contribution from the insect body (I_B) from the diffuse and specular wing reflections (I_{DW}, I_{SW}). Using these signals, we could estimate morphological and behavioural properties.

The calculated features used in this study are presented in Table 1 and a formal mathematical formulation of the features extraction algorithms are included as supplementary material. An overview of the median and inter quantile range (IQR) of the extracted features from individual insect observations, aggregated by family level, is presented in Figure 3.

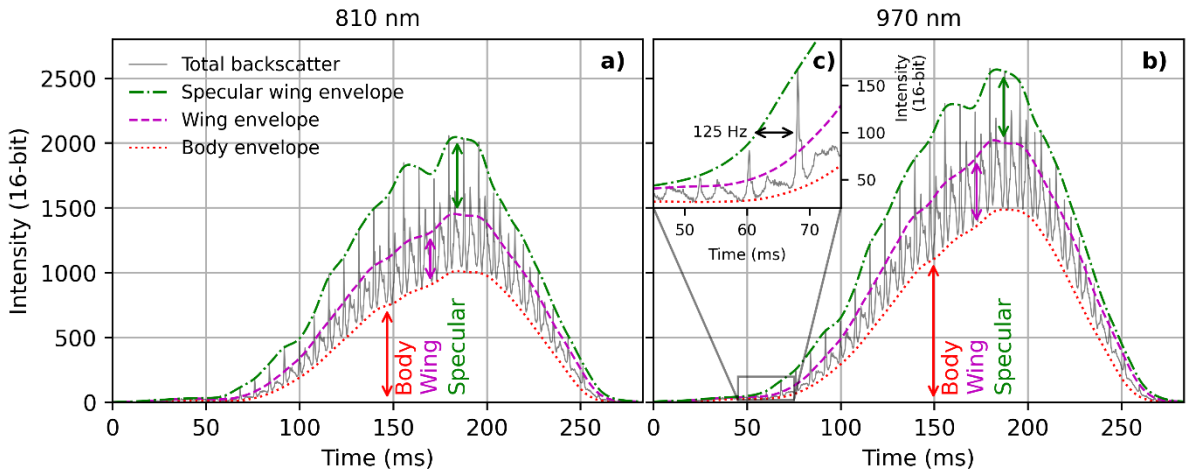


Figure 2: *Example of an insect recording of a solitary bee (Andrenidae vaga) by the sensor. Wingbeats are visible as a modulation on the signal. The body-, wing- and specular contributions are isolated by moving average and maximum filters. In this case, the insect body has a higher reflectance at 970 nm in b) than for 808 nm in a), while the amplitude of the wing signal is similar. This indicates a melanised body. A zoomed in view of the specular and diffuse wing envelopes is shown in c).*

Table 1. *Calculated features from insect observations. All features, except the wing beat frequency, are defined as ratios between zero and one.*

Feature	Variable	Values	Unit	Definition	Ref
Wing beat frequency	WBF	20..1000	Hz		[39]–[42]
Body to wing ratio	BWR	0..1	-	$I_{B,}/(I_B + I_{DW})$	[39], [43]
Specular to diffuse wing ratio	SWR	0..1	-	$I_{SW}/(I_{DW} + I_{SW})$	[39], [43]
Body to specular wing	BSR	0..1	-	$I_B/(I_B + I_{SW})$	[39], [43]
Body melanization	BM	0..1	-	$I_{B\text{-}970}/(I_{B\text{-}808}+ I_{B\text{-}970})$	[39], [43]
Wing melanization	WM	0..1	-	$I_{DW\text{-}970}/(I_{DW\text{-}808}+I_{DW\text{-}970})$	[39], [43]
Specular ratio	SR	0..1	-	$I_{SW\text{-}808}/(I_{SW\text{-}808} + I_{SW\text{-}970})$	[39], [43]

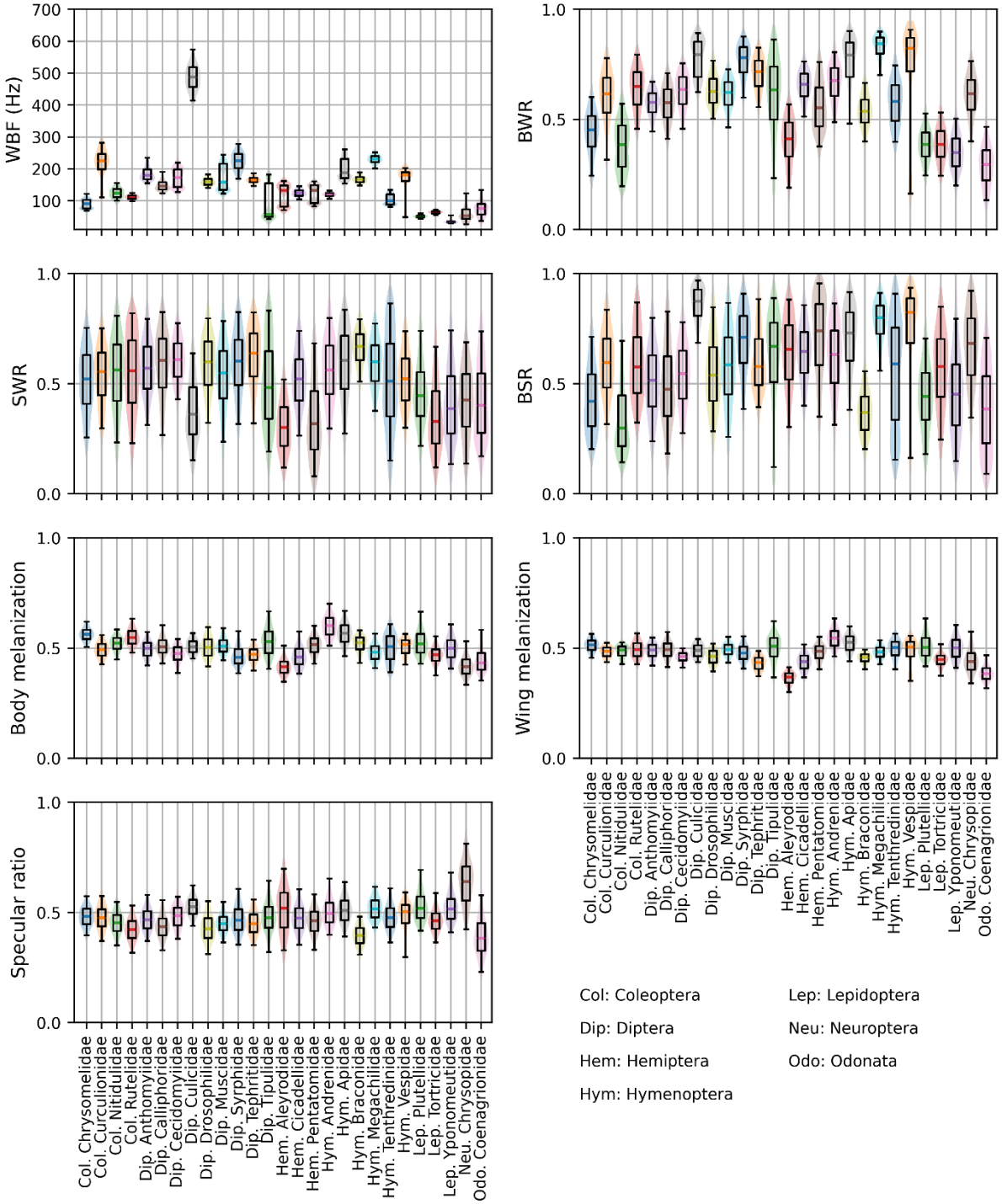


Figure 3: *Calculated features from the single species insect data, aggregated per family. Horizontal line indicates the median, boxes the 25-75% IQR and whiskers the 5-95% IQR. A shaded violin plot is drawn behind the boxes.*

3.2 Oscillation power spectra

As a high dimensional complement to the low dimensional feature representation of the time signals, we calculate the oscillatory power spectra for each insect observation. These oscillation

powers and overtones are estimated by [44] and are contributed by the body, wing movements and specular reflections of glossy wings [30]. The calculation result in 100 frequency bins from 0 to 2 kHz with 20 Hz steps for both wavelengths. The modulation spectra were subsequently concatenated and auto-normalized. The median- and IQR power spectra are presented for each family in Figure 4.

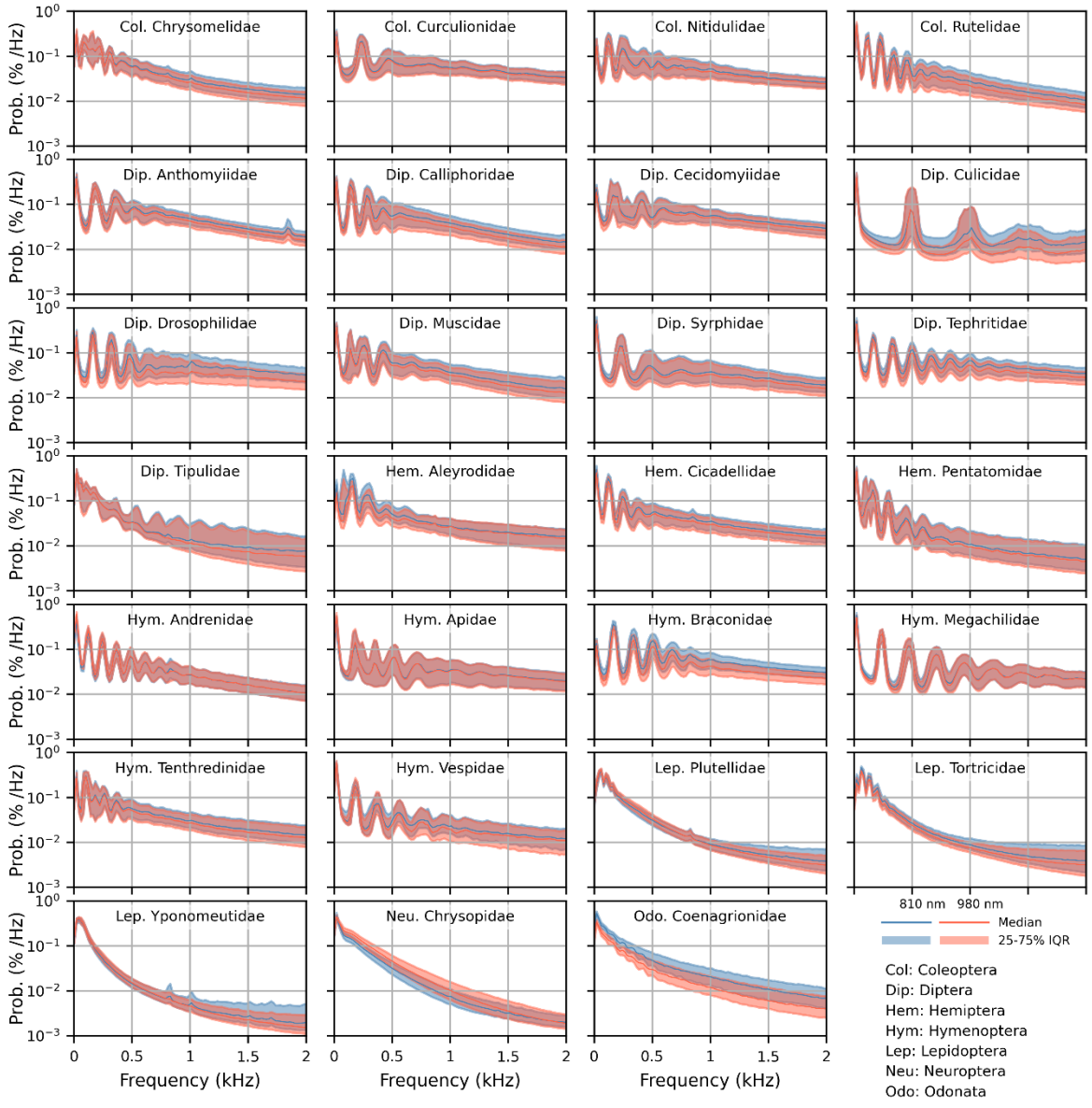


Figure 4: *Frequency power spectra from the caged single species recordings aggregated on family level. The median is plotted as a solid line and the 25-75% IQR as a shaded area. The zero Hz intercept correspond to body size, and the first peak is the fundamental wingbeat frequency. Following peaks are the overtones and a large number of overtones correspond to sharp, specular reflexes from the wings. For example, Hymenoptera Megachilidae have a wingbeat frequency around 230 Hz and several overtones at 460, 690 and 920 Hz.*

4. Clustering insect data

We evaluated two unsupervised clustering methods, with the aim of using the estimated number of clusters as an indicator for species richness. The first method was Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [45] and the second method was Hierarchical Clustering Analysis (HCA) [46].

In the DBSCAN algorithm, a random insect observation is selected as a starting point and all data points within a certain radius, ε, in the parameter space are added into its cluster. If less than *MinPts* data points are found within ε, the observation is discarded as noise and the algorithm continues with a randomly selected new observation until all data points are treated. The time signal features shown in figure 3 were mean scale normalized and clustered using DBSCAN and an estimated number of clusters, R_{DBSCAN}, is obtained. We used the scikit-learn package implementation in Python [47].

In HCA, all observations starts as individual clusters. The pairwise Euclidian distances in the parameter space, and the corresponding linkage vector, between all insect observations are calculated and used to merge all clusters sequentially until only a single cluster remains. The linkages were calculated using Ward's method [46] which merges clusters by minimizing the variance in each cluster. Pairwise distances and linkage vectors were calculated using the scikit-learn and fastcluster packages in Python [48]. HCA has previously been used to cluster insect observations in entomological lidar and to estimate the number of insect groups [34], [35]. In similar fashion, we applied the HCA to the logarithmized oscillatory power spectra. The linkage vectors contain information on the diversity of power spectra various methods can be applied to determine the appropriate number of clusters from it [49]–[51]. As the previously used elbow method [49] performed poorly, we developed a novel threshold according to:

$$R_{HCA} = \sum \left(\frac{L_n}{\sqrt{N}} > p_0 n^{p_1} \right) \quad (1)$$

where L_n contains the linkages, N is the number of observations, $n \in \{1 \ldots N\text{-}1\}$, p_0 and p_1 are tuning parameters.

4.1 Performance tests on flight cage data

During method development we defined two tests to evaluate the performance of the two clustering algorithms.

Richness test

The first test investigates the correlation between the estimated number of clusters (R_{DBSCAN}, R_{HCA}) and true number of included species, R_{spec}. We generated a selection of 1000 observations evenly distributed among $R_{spec} \in \{1 \ldots 42\}$, by random selection without replacement. The observations were fed into the two algorithms and the number of clusters calculated. The correlation between the logarithmized number of clusters, R_{DBSCAN}, R_{HCA}, and R_{spec} was calculated by linear regression where

$$R_{spec} = \beta_0 R_{clust} + \beta_1 \quad (2)$$

and the corresponding Pearson correlation ρ was measured. To improve the dynamic range and punish configurations that yielded a very low number of clusters, despite yielding high values of ρ, the performance metric Q_{rich} was defined as

$$Q_{rich} = \rho(R_{spec}, R_{clust})\ \beta_0 \quad (3)$$

Abundance insensitivity test

The second test investigates whether the number of found clusters R_{clust} was related to the number of included insect recordings, N_{obs}. A new randomized dataset was created by randomly selecting one species and linearly increasing the number of insect observations from 100 to 1000 in 20 steps. As R_{spec} was kept at one species during the entire test, the quality metric was defined as

$$Q_{flat} = \mu(R_{clust})/\sigma(R_{clust}) \quad (4)$$

Where μ and σ is the mean and standard deviation respectively.

4.2 Clustering parameter tuning

To compare the performance of both tests, a quality criterion Q was defined as

$$Q = Q_{rich} * Q_{flat} \quad (5)$$

Where higher is better and Q_{rich} and Q_{flat} were truncated at 0 to remove any negative correlations. The tests were conducted on random samples of the data, as described below. Each test was repeated on 50 randomized permutations for both methods. The best performing parameters were found by a grid search at 450 points for DBSCAN and 314 points for HCA and used in the field evaluation.

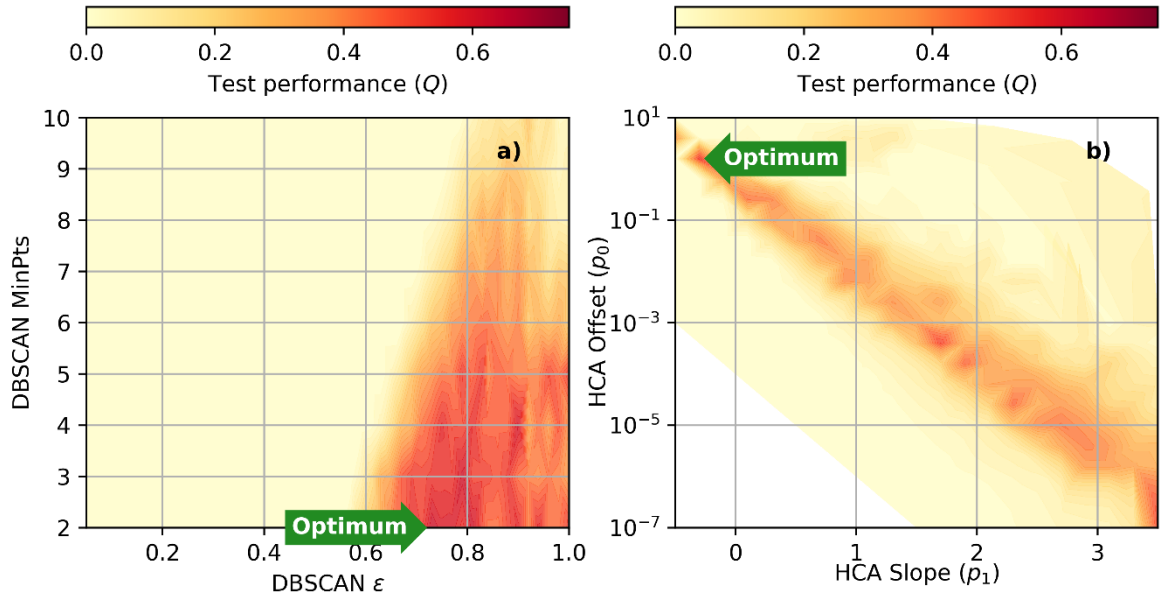


Figure 5: *Optimization landscape for the two clustering algorithms. Both methods find a clearly defined optimum.*

4.3 Field validation

The best performing configurations of the DBSCAN and HCA methods on the flight cage data were applied to the data recorded in the field. Insect observations from each sensor were aggregated by the weekly emptying schedule of the corresponding Malaise trap. The aggregated data was fed to the clustering algorithms and the number of found clusters was compared to the number of families identified in the corresponding Malaise traps. Due to the non-normal distribution of the data, the number of clusters and family richness were logarithmized before calculating the Pearson's correlation.

5. Results

5.1. Labelled data tests

The optimization landscapes for both methods are presented in Figure 5. The best performing parameters were ε=0.67, *MinPts* = 2 for the DBSCAN method and p_0=1.58, p_1 = -0.3 for the HCA. The individual test results for the best performing parameters are presented in Figure 6.

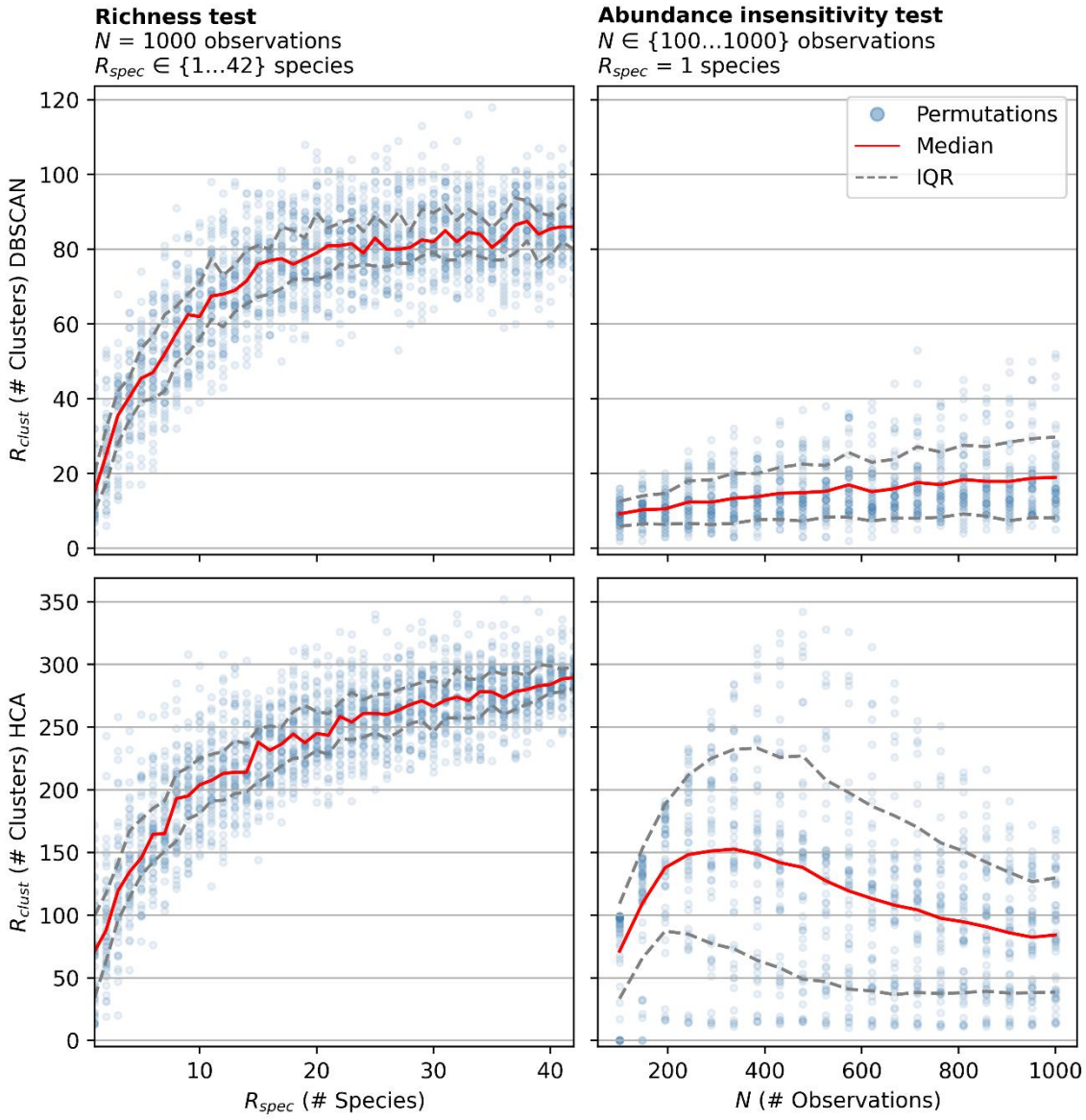


Figure 6: *The results of the richness and abundance insensitivity tests for the best performing parameters of the DBSCAN (upper) and HCA (lower) methods. Individual clustering permutations are plotted as dots. The median is plotted in red and 25, 75% quantiles in dashed grey. In the richness test (left column), each dot represents a dataset with 1000 insect observations and the ideal result is a linear relationship between* R_{spec} *and* R_{clust}*. In the abundance insensitivity test (right column), each dot represents a dataset from a single species and the ideal result is a constant, low number of clusters.*

5.2. Field validation

After sub-sampling the Malaise trap catch samples, a total of 9503 insects from 11 orders and 195 families were morphologically identified. The most prevalent orders were Diptera and Hymenoptera, accounting for 75% and 15% of all identified insects, respectively. Among the families, *Diptera Sciaridae sp*. was the most common, accounting for 25% of all identified insects. Figure 9 shows the sample times and abundance in each trap .

The sensors were active for a total of 75 weeks and recorded 2,402,345 observations. After removing observations caused by rain and dust (2,113,737) or where the feature extraction algorithms failed to estimate features (50,024), 238,584 observations remained available for clustering.

Table 2. *Overview of field samples for Malaise traps and sensors.*

Biotope	*Measurement period*	*Location*	*Collected trap samples*	*Identified trap samples*	*No. Weeks Sensors*	*Used in comparison*
Spruce forest	16/3-20/7 & 3/8-2/11	56°20'08.1"N 14°23'09.4"E	28	20	27	16
Deciduous forest	8/6 - 9/11	55°41'46.4"N 13°26'51.5"E	18	15	20	12
Grazed grassland 1	1/4 - 1/7	55°49'40.4"N 11°26'10.7"E	10	2	13	1
Grazed grassland 2	8/4 – 10/6	55°48'07.0"N 11°23'18.1"E	17	8	8	6
Oilseed rape	1/4 – 27/5	55°29'03.1"N 11°29'32.6"E	5	5	7	5

For the 40 weeks were both sensor and trap data were available, the family richness in the Malaise traps was correlated with the number of clusters. Due to the non-normal distribution of the data, all measures were logarithmized. The Pearson correlation show that the sensor methods explain close to 40% of the variation in time and space of abundance and 30 to 45% of that of diversity (Figure 7, Table 3). The correlation between abundance and richness estimates is strong for all methods, including the traps, but especially for the HCA method.

Table 3. *Pearson correlations between logarithmized biodiversity estimates. Calculated for the weeks were both trap and sensor data were available. All correlations are significant with $p < 0.05$.*

	N_{trap}	R_{trap}	N_{sensor}	R_{DBSCAN}	R_{HCA}
N_{trap}	1	0.73	0.55	0.54	0.63
R_{trap}	0.73	1	0.59	0.54	0.67
N_{sensor}	0.55	0.59	1	0.69	0.97
R_{DBSCAN}	0.54	0.54	0.69	1	0.75
R_{HCA}	0.63	0.67	0.97	0.75	1

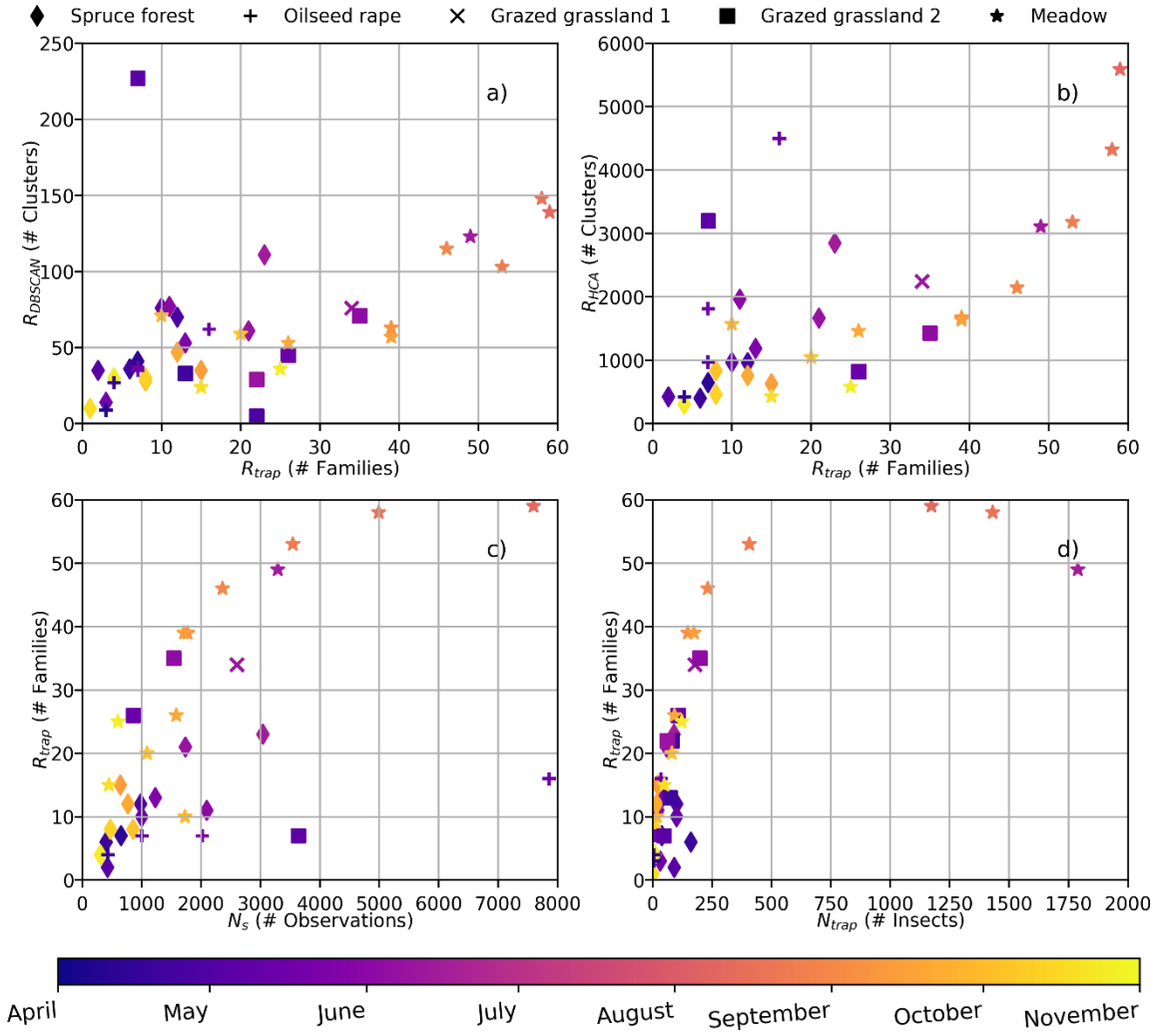


Figure 7. *Relationship between measured and estimated abundance and biodiversity. a) The number of clusters found by the DBSCAN vs. the trap richness. b) The number of clusters found by the HCA vs- the trap richness. c) The trap richness vs the number of insect observations recorded by the sensor. d) The richness vs the number of caught insects in the Malaise traps.*

Comparing insect abundance and richness evaluated by traps and sensors over the full experiment period, the sensors in general yield one order of magnitude more observations per week. However, the reported number of insects from the traps are sub-sampled by a factor 6. Both sensors and traps find the highest biodiversity and insect abundance at the meadow site during summer as shown in figure 8.

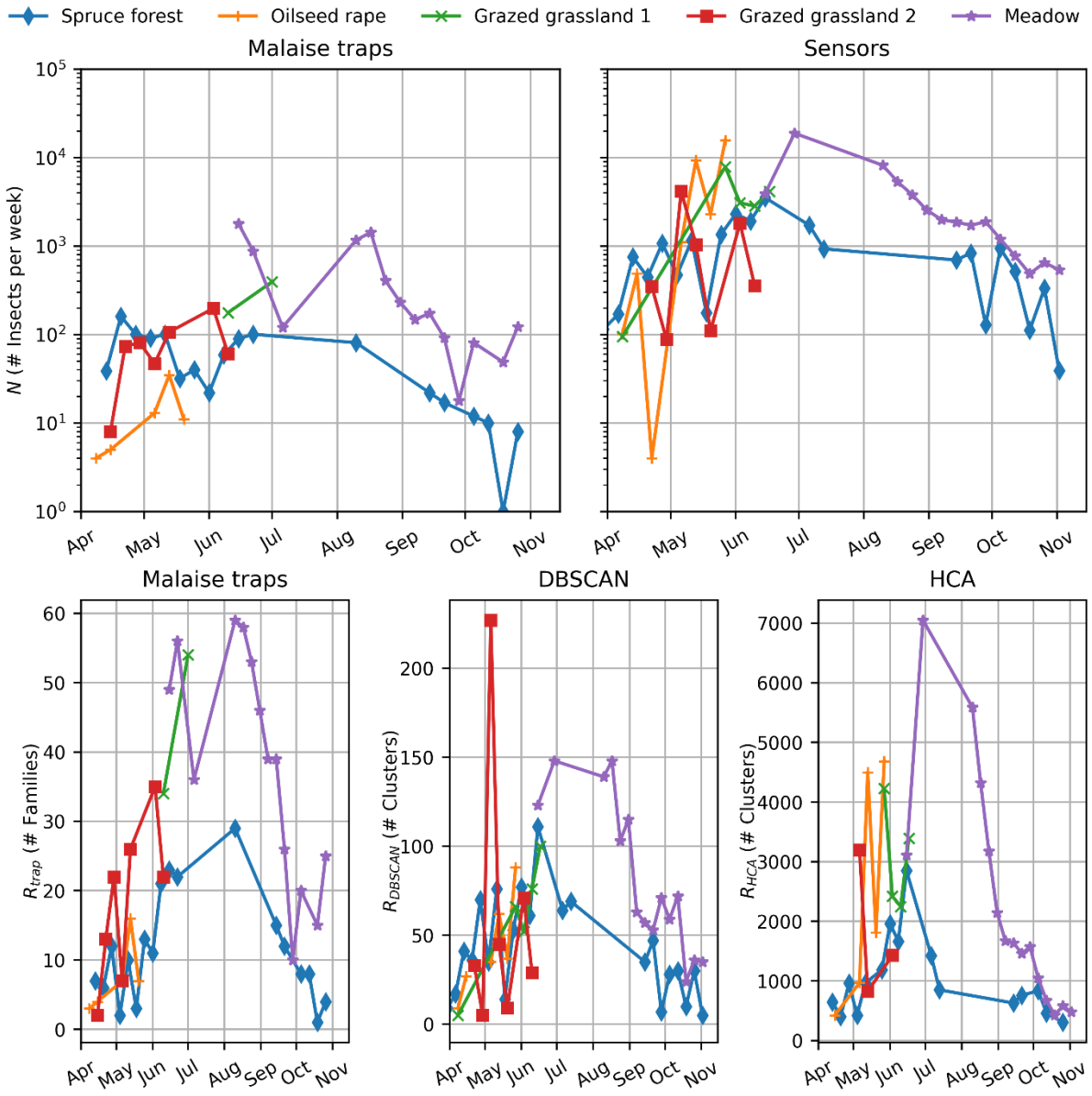


Figure 8: *Sensor and Malaise trap data from the full measurement period. Upper figures show the number of insects per week in Malaise traps (left) and sensors (right). Lower row show the measured family richness in the Malaise traps, and the number of clusters found by the algorithms. Due to hardware and connectivity issues, some data points are missing from the sensor data.*

6. Discussion

Both unsupervised clustering methods of categorizing optically recorded insect signals correlate well with estimates obtained by Malaise traps. This suggests that the optical sensors and similar instrumentation are viable tools for monitoring not only insect abundance, but also richness. Both algorithms yielded higher numbers of clusters than the measured family richness in the traps, but that may reflect that our clustering method had a higher resolution than our identification to family level of Malaise trap catches [52]. However, the match was not perfect, with correlations of 54% and 67%, which may be caused by how both the photonic and the Malaise trap methods are both estimates of the true biodiversity of insects [53].

While correlated, it is difficult to relate the number of clusters to an absolute number of species in the field. Many species have distinct differences in wing beat frequency between sexes [39]. It is also known from literature that even a single species and sex can produce distinct optical signals depending on observation aspect [39], [54]. Therefore, photonic insect recordings can be challenging to differentiate due to overlapping features in the parameter space [39], [43], [55], [56]. While various machine learning methods still can perform very well, any photonic insect sensor will ultimately have a limit for the distinguishable number of species, depending on the instrument complexity. Despite these arguments, there is an expectation that a more diverse composition of species will produce a more diverse composition of optical signals, which is confirmed in this work.

The number of observations recorded in the field was generally higher than the number of observations included in the caged data tests. We believe this raises some noticeable differences. The abundance insensitivity test was designed to minimize the correlation between the number of observations and the number of clusters. However, as seen in figure 6, In this test, the DBSCAN show a slow increase in the number of clusters while the HCA shows the opposite. However, on the field data the HCA is 97% correlated with the number of observations. We believe that a tuning of the parameters optimized on the caged data using data from the field would reduce this correlation but to reduce the risk of overfitting, we have avoided such adjustments in this work.

In the field data, the number of observations recorded by the sensor is more correlated with the trap richness than the estimated richness by the DBSCAN. While the richness and abundance also were strongly correlated in the Malaise traps we do not believe abundance can be used as a biodiversity measure alone as this study primarily included protected areas where the insect abundance and richness were strongly correlated.

The number of observations in each cluster were not well correlated with the distribution of families in the traps. It is therefore difficult to compare measures such as Shannon's or Simpson's index between the clustering methods and traps. The correlation between the sensors and traps of these indices are included as supplementary material.

Comparing the methodology of all methods, the task of richness estimation closely resembles the challenge of finding and counting rare outliers. In identification of trap catches, single species observations could account for up to 30% of the identified richness [57], [58]. As DBSCAN cannot support clusters with single members, a larger number of observations per species are required in order to provide a meaningful estimate. In future work, the number of

singletons could be included in the richness estimate, but as the number of singletons can be large, there is a risk that they would outweigh the number of clusters and drive the entire metric.

7. Outlook and conclusions

Photonic methods have a high potential to benefit applied insect science and management for several reasons. First, it enables cost-efficient collection of data with a resolution in time and space that is hard to achieve using manual methods based on catches. This makes it an ideal method both for long-term monitoring and for detecting consequences of management interventions such as use of pesticides where local effects may be transient yet consequential. Second, it allows for monitoring inside protected areas, where destructive sampling may be non-ethical or legally prohibited. However, as a method, its value will be related to its accuracy and taxonomic resolution, while further development and quality controls are warranted.

In this work, we have shown that photonic sensors in combination with unsupervised clustering algorithms are able to measure the species richness *in situ*. We tested two different algorithms, DBSCAN and HCA, on a low and high dimensional representation of the insect recordings respectively. Both methods showed a correlation between the number of clusters and the species richness in flight cage experiments. Further, comparison to diversity measures determined from conventional trapping yielded good correlation between sensors and Malaise traps richness on family level. This shows that photonic sensors can successfully be used for monitoring not only insect abundance but also diversity.

8. Acknowledgements

Hjalte Ro-Poulsen, Anna Runemark, Emma Kärrnäs, Meng Li, Maja Olofsson assisted with collection of Malaise trap samples. Miljøstyrelsen Denmark, Rachel Muheim and Bengt Augustsson provided access to field sites. A special thanks to Jesper Lemmich for enabling taxonomic identification.

9. Conflicts of interest

Klas Rydhmer (KR), Samuel Jansson (SJ), Laurence Still (LS), Brittany D. Beck (BB), Vasileia Chatzaki (VC), Karen Olsen (KO), Bennett Van Hoff (BH), Christoffer Grønne (CG), Jakob Klinge Meier (JM), Marta Montaro (MM), Carsten Kirkeby (CK), and Mikkel Brydegaard (MB) are current or previous employees and/or shareholders in FaunaPhotonics A/S who produced the sensors used in this study. We declare that this has not impacted the work in any way.

10. Author Contributions

KR designed the study, conducted the field trials, conducted the data analysis and wrote the first version of the manuscript. SJ and LS contributed to early development of the DBSCAN algorithm and editing of the manuscript. BB, VC, JM, KO, BH, CG, JM and MM conducted the taxonomic identification of the Malaise trap catches. CK, Inger Kappel Schmidt (IKS) contributed to experiment design and editing of the manuscript. Henrik Smith (HS) contributed to experiment design and writing of the manuscript. MB contributed to experiment design, data analysis, algorithm development and writing of the manuscript. All authors read and approved the final version of the manuscript.

11. Data availability

An online repository with all data needed to recreate the figures in this manuscript will be made available upon paper acceptance.

12. References

[1] C. A. Hallmann *et al.*, "More than 75 percent decline over 27 years in total flying insect biomass in protected areas," *PLoS One*, vol. 12, no. 10, 2017, doi: 10.1371/journal.pone.0185809.

[2] R. van K. and D. E. B. and K. B. G. and A. B. S. and A. G. and J. M., "Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances," *Science (1979)*, vol. 368, no. 6489, pp. 417–420, 2020, doi: 10.1126/science.abd8947.

[3] S. Seibold *et al.*, *Arthropod decline in grasslands and forests is associated with landscape-level drivers*, vol. 574, no. 7780. 2019. doi: 10.1038/s41586-019-1684-3.

[4] R. van Klink *et al.*, "Emerging technologies revolutionise insect ecology and monitoring," *Trends Ecol Evol*, vol. 37, no. 10, pp. 872–885, 2022, doi: 10.1016/j.tree.2022.06.001.

[5] F. Neff *et al.*, "Different roles of concurring climate and regional land-use changes in past 40 years' insect trends," *Nat Commun*, vol. 13, no. 1, p. 7611, 2022, doi: 10.1038/s41467-022-35223-3.

[6] D. L. Wagner, E. M. Grames, M. L. Forister, M. R. Berenbaum, and D. Stopak, "Insect decline in the Anthropocene: Death by a thousand cuts," *Proc Natl Acad Sci U S A*, vol. 118, no. 2, Jan. 2021, doi: 10.1073/PNAS.2023989118.

[7] I. I. R. B. Y. UN, "Convention on biological diversity," *Treaty Collection*, 1992.

[8] W. W. Weisser and E. Siemann, *Insects and ecosystem function*, vol. 173. Springer Science & Business Media, 2013.

[9] J. E. LOSEY and M. VAUGHAN, "The Economic Value of Ecological Services Provided by Insects," *Bioscience*, vol. 56, no. 4, p. 311, 2006, doi: 10.1641/0006-3568(2006)56[311:tevoes]2.0.co;2.

[10] M. Loreau, A. Hector, and F. Isbell, *The ecological and societal consequences of biodiversity loss*. John Wiley & Sons, 2022.

[11] M. Dainese *et al.*, "A global synthesis reveals biodiversity-mediated benefits for crop production," *Georg K. S. Andersson*, vol. 19, Accessed: Mar. 08, 2023. [Online]. Available: http://advances.sciencemag.org/

[12] E. C. Oerke, "Crop losses to pests," *J Agric Sci*, vol. 144, no. 1, pp. 31–43, Feb. 2006, doi: 10.1017/S0021859605005708.

[13] D. Karlsson, E. Hartop, M. Forshage, M. Jaschhof, and F. Ronquist, "The Swedish Malaise Trap Project: A 15 Year Retrospective on a Countrywide Insect Inventory," *Biodivers Data J*, vol. 8, p. e47255, Jan. 21AD, [Online]. Available: https://doi.org/10.3897/BDJ.8.e47255

[14] "About - The SPRING project - Strengthening Pollinator Recovery through INdicators and monitorinG." https://www.ufz.de/spring-pollination/index.php?en=49075 (accessed Mar. 08, 2023).

[15] L. Wührl *et al.*, "DiversityScanner: Robotic handling of small invertebrates with machine learning methods," *Mol Ecol Resour*, vol. 22, no. 4, pp. 1626–1638, 2022, doi: 10.1111/1755-0998.13567.

[16] J. Ärje *et al.*, "Automatic image-based identification and biomass estimation of invertebrates," *Methods Ecol Evol*, vol. 11, no. 8, pp. 922–931, 2020, doi: 10.1111/2041-210X.13428.

[17] P. Taberlet, E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev, "Towards next-generation biodiversity assessment using DNA metabarcoding," *Mol Ecol*, vol. 21, no. 8, pp. 2045–2050, Apr. 2012, doi: 10.1111/J.1365-294X.2012.05470.X.

[18] C. S. Svenningsen *et al.*, "Detecting flying insects using car nets and DNA metabarcoding," *Biol Lett*, vol. 17, no. 3, Mar. 2021, doi: 10.1098/RSBL.2020.0833.

[19] P. D. Lamb, E. Hunter, J. K. Pinnegar, S. Creer, R. G. Davies, and M. I. Taylor, "How quantitative is metabarcoding: A meta-analytical approach," *Mol Ecol*, vol. 28, no. 2, pp. 420–430, Jan. 2019, doi: 10.1111/MEC.14920.

[20] A. O. Shelton *et al.*, "Toward quantitative metabarcoding," *Ecology*, vol. 104, no. 2, p. e3906, Feb. 2023, doi: 10.1002/ECY.3906.

[21] P. J. Stephenson *et al.*, "Measuring the Impact of Conservation: The Growing Importance of Monitoring Fauna, Flora and Funga," *Diversity 2022, Vol. 14, Page 824*, vol. 14, no. 10, p. 824, Sep. 2022, doi: 10.3390/D14100824.

[22] M. Besson *et al.*, "Towards the fully automated monitoring of ecological communities," *Ecol Lett*, vol. 25, no. 12, pp. 2753–2775, Dec. 2022, doi: 10.1111/ELE.14123.

[23] I. Potamitis, P. Eliopoulos, and I. Rigakis, "Automated Remote Insect Surveillance at a Global Scale and the Internet of Things," *Robotics*, vol. 6, no. 3, 2017, doi: 10.3390/robotics6030019.

[24] D. F. Silva, V. M. A. De Souza, K. E. Batista GEAPA, and D. P. W. Ellis, “Applying machine learning and audio analysis techniques to insect recognition in intelligent traps. Proceedings—2013 12th International Conference on Machine Learning and Applications, ICMLA 2013. 2013.” 2013.

[25] M. E. Sinka *et al.*, “HumBug – An Acoustic Mosquito Monitoring Tool for use on budget smartphones,” *Methods Ecol Evol*, vol. 12, no. 10, pp. 1848–1859, Oct. 2021, doi: 10.1111/2041-210X.13663.

[26] A. Jeliazkov, Y. Bas, C. Kerbiriou, J. F. Julien, C. Penone, and I. Le Viol, “Large-scale semi-automated acoustic monitoring allows to detect temporal decline of bush-crickets,” *Glob Ecol Conserv*, vol. 6, pp. 208–218, Apr. 2016, doi: 10.1016/J.GECCO.2016.02.008.

[27] O. L. P. Hansen *et al.*, “Species-level image classification with convolutional neural network enables insect identification from habitus images,” *Ecol Evol*, 2019, doi: 10.1002/ece3.5921.

[28] S. Butail, N. Manoukis, M. Diallo, J. M. Ribeiro, T. Lehmann, and D. A. Paley, “Reconstructing the flight kinematics of swarming and mating in wild mosquitoes,” *J R Soc Interface*, vol. 9, no. 75, pp. 2624–2638, Oct. 2012, doi: 10.1098/RSIF.2012.0150.

[29] K. Rydhmer *et al.*, “Automating insect monitoring using unsupervised near-infrared sensors,” *Sci Rep*, vol. 12, no. 1, pp. 1–11, 2022.

[30] M. Brydegaard and S. Svanberg, “Photonic monitoring of atmospheric and aquatic fauna,” *Laser Photon Rev*, vol. 12, no. 12, p. 1800135, 2018.

[31] M. J. Tauc, K. M. Fristrup, K. S. Repasky, and J. A. Shaw, “Field demonstration of a wing-beat modulation lidar for the 3D mapping of flying insects,” *OSA Contin*, vol. 2, no. 2, p. 332, 2019, doi: 10.1364/osac.2.000332.

[32] A. P. Genoud, T. Saha, G. M. Williams, and B. P. Thomas, “Insect biomass density: measurement of seasonal and daily variations using an entomological optical sensor,” *Appl Phys B*, vol. 129, no. 2, pp. 1–13, 2023, doi: 10.1007/s00340-023-07973-5.

[33] A. P. Genoud, G. M. Williams, and B. P. Thomas, “Continuous monitoring of aerial density and circadian rhythms of flying insects in a semi-urban environment,” *PLoS One*, vol. 16, no. 11, p. e0260167, Nov. 2021, [Online]. Available: https://doi.org/10.1371/journal.pone.0260167

[34] M. Brydegaard *et al.*, “Lidar reveals activity anomaly of malaria vectors during pan-African eclipse,” *Sci Adv*, vol. 6, May 2020, doi: 10.1126/sciadv.aay5487.

[35] B. K. Kouakou, S. Jansson, M. Brydegaard, and J. T. Zoueu, “Entomological Scheimpflug lidar for estimating unique insect classes in-situ field test from Ivory Coast,” *OSA Contin*, vol. 3, no. 9, pp. 2362–2371, Sep. 2020, doi: 10.1364/OSAC.387727.

[36] H. Ro-Poulsen, “Wild bees and honey bees in semi-natural habitats of Denmark - Effects of beekeeping and habitat management in grasslands and heathlands,” University of Copenhagen, 2023.

[37] M. Li *et al.*, “Bark beetles as lidar targets and prospects of photonic surveillance,” *J Biophotonics*, no. November, pp. 1–16, 2020, doi: 10.1002/jbio.202000420.

[38] C. Kirkeby, M. Wellenreuther, and M. Brydegaard, “Observations of movement dynamics of flying insects using high resolution lidar,” *Sci Rep*, vol. 6, 2016, doi: 10.1038/srep29083.

[39] A. Gebru, S. Jansson, R. Ignell, C. Kirkeby, J. C. Prangsma, and M. Brydegaard, “Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight,” *J Biophotonics*, vol. 11, no. 8, 2018, doi: 10.1002/jbio.201800014.

[40] I. Potamitis and I. Rigakis, “Measuring the fundamental frequency and the harmonic properties of the wingbeat of a large number of mosquitoes in flight using 2D optoacoustic sensors,” *Applied Acoustics*, vol. 109, pp. 54–60, Aug. 2016, doi: 10.1016/j.apacoust.2016.03.005.

[41] A. Moore, J. R. Miller, B. E. Tabashnik, and S. H. Gage, “Automated identification of flying insects by analysis of wingbeat frequencies,” *J Econ Entomol*, vol. 79, no. 6, pp. 1703–1706, 1986.

[42] E. Malmqvist, S. Jansson, S. Török, and M. Brydegaard, “Effective parameterization of laser radar observations of atmospheric fauna,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, p. 1, 2015, doi: 10.1109/JSTQE.2015.2506616.

[43] C. Kirkeby *et al.*, “Advances in automatic identification of flying insects using optical sensors and machine learning,” *Sci Rep*, vol. 11, no. 1, p. 1555, 2021, doi: 10.1038/s41598-021-81005-0.

[44] P. D. Welch, “The Use of Fast Fourier Transform for the Estimation of Power Spectra,” *Digit Signal Process*, no. 2, pp. 532–574, 1975.

[45] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise.,” in *kdd*, 1996, pp. 226–231.

[46] J. H. Ward, “Hierarchical Grouping to Optimize an Objective Function,” *J Am Stat Assoc*, vol. 58, no. 301, pp. 236–244, 1963, doi: 10.1080/01621459.1963.10500845.

[47] F. Pedregosa FABIANPEDREGOSA *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[48] D. Müllner, “fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python,” *J Stat Softw*, vol. 53, no. 9, pp. 1–18, May 2013, doi: 10.18637/JSS.V053.I09.

[49] R. L. Thorndike, “Who belongs in the family?,” *Psychometrika*, vol. 18, no. 4, pp. 267–276, Dec. 1953, doi: 10.1007/BF02289263/METRICS.

[50] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *J R Stat Soc Series B Stat Methodol*, vol. 63, no. 2, pp. 411–423, Jan. 2001, doi: 10.1111/1467-9868.00293.

[51] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J Comput Appl Math*, vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[52] A. Báldi, “Using higher taxa as surrogates of species richness: A study based on 3700 Coleoptera, Diptera, and Acari species in Central-Hungarian reserves,” *Basic Appl Ecol*, vol. 4, no. 6, pp. 589–593, 2003, doi: 10.1078/1439-1791-00193.

[53] J. Uhler *et al.*, “A comparison of different Malaise trap types,” 2022, doi: 10.1111/icad.12604.

[54] M. Li *et al.*, “Bark beetles as lidar targets and prospects of photonic surveillance,” *J Biophotonics*, no. January 2021, 2020, doi: 10.1002/jbio.202000420.

[55] A. P. Genoud, Y. Gao, G. M. Williams, and B. P. Thomas, “A comparison of supervised machine learning algorithms for mosquito identification from backscattered optical signals,” *Ecol Inform*, vol. 58, p. 101090, Jul. 2020, doi: 10.1016/j.ecoinf.2020.101090.

[56] Y. Chen, A. Why, G. Batista, A. Mafra-Neto, and E. Keogh, “Flying insect classification with inexpensive sensors,” *J Insect Behav*, vol. 27, no. 5, pp. 657–677, 2014.

[57] O. H. Diserud, E. Stur, and K. Aagaard, “How reliable are Malaise traps for biomonitoring? - A bivariate species abundance model evaluation using alpine Chironomidae (Diptera),” *Insect Conserv Divers*, vol. 6, no. 5, pp. 561–571, 2013, doi: 10.1111/icad.12012.

[58] F. W. Preston, “The Commonness, And Rarity, of Species,” *Ecology*, vol. 29, no. 3, pp. 254–283, Jul. 1948, doi: 10.2307/1930989.

Appendix A – Data processing and feature extraction

Insect observations are recorded as dual band signals in four segments, yielding a time signal with eight channels. Each segment covers a different volume in front of the sensor. To reduce artefacts from very weak signals, each channel is filtered individually, and empty channels are removed from further analysis.

The filtering consists of an intensity threshold and a classification step. The intensity thresholding asserts that channel has a maximum intensity above 20 counts. The classification step applies a proprietary noise classifier. Unless an insect signal is observed in all segments, the channels from empty segments are discarded. We believe the design of this classifier is out of scope for this work, but it can be summarized as a one-dimensional convolutional neural network trained on manually annotated data from both cages and the field.

The insect body, diffuse and specular wing contributions was estimated from the time signal $I(t, \lambda)$ by running a moving minimum, average and maximum filter over the signal in each channel, using the extracted wingbeat period as filter width according to

$$\delta t = \frac{1}{WBF}$$

$$I_B(t,\lambda) = \|I(t,\lambda) \dots I(t+\delta t,\lambda)\|_{min} \otimes e^{-\frac{(t-T)^2}{\delta t^2}}$$

$$I_{DW}(t,\lambda) = \|I(t,\lambda) \dots I(t+\delta t,\lambda)\|_{med} \otimes e^{-\frac{(t-T)^2}{\delta t^2}}$$

$$I_{SW}(t,\lambda) = \|I(t,\lambda) \dots I(t+\delta t,\lambda)\|_{max} \otimes e^{-\frac{(t-T)^2}{\delta t^2}}$$

To cut away the signal rising and falling flanks of the signal, a centre time, t_0, and transit time, Δt, was calculated accordingly:

$$t_0 = \frac{\sum_t t \sum_\lambda I(t,\lambda)}{\sum_t \sum_\lambda I(t,\lambda)}$$

$$\Delta t = \sqrt{\frac{\sum_t (t-t_0)^2 \sum_\lambda I(t,\lambda)}{\sum_t \sum_\lambda I(t,\lambda)}}$$

Scalar estimates were obtained by calculating median values within a time window of $t_0 \pm \Delta t$:

$$\overline{I_B}(\lambda) = \|I_B(t_0 - \Delta t \dots t_0 + \Delta t,\lambda)\|_{med}$$

$$\overline{I_{DW}}(\lambda) = \|I_{DW}(t_0 - \Delta t..t_0 + \Delta t,\lambda)\|_{med} - \overline{I_B}(\lambda)$$

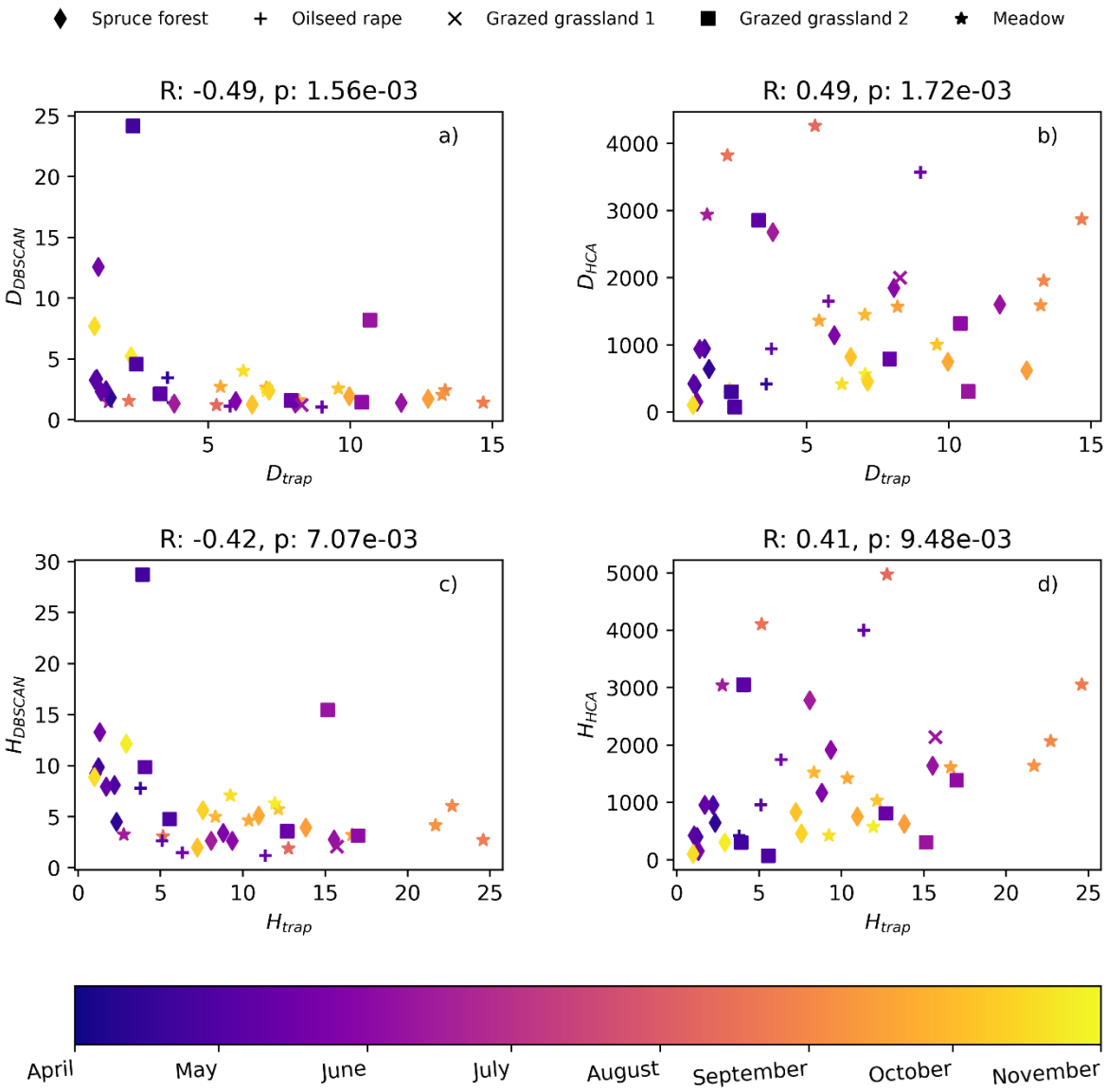
$$\overline{I_{SW}}(\lambda) = \|I_{SW}(t_0 - \Delta t..t_0 + \Delta t,\lambda)\|_{med} - \overline{I_B}(\lambda)$$

Calculated features are averaged over all non-empty channels. Dual band features (body and wing melanization and specular ratio) are averaged over all non-empty segments. For single band features (BWR, SWR, BSR) we chose the features recorded in the 808 nm band to limit the number of features used in the model. For the wingbeat frequency, we calculated the median across all non-zero channels.

Appendix B – Labelled data species:

Order	Family	Species	Number of observations
Coleoptera	Chrysomelidae	alni	1310
Coleoptera	Chrysomelidae	chrysocephalus	1564
Coleoptera	Curculionidae	pallidactylus	1277
Coleoptera	Nitidulidae	aeneus	2678
Coleoptera	Rutelidae	japonica	3674
Diptera	Anthomyiidae	antiqua	1733
Diptera	Calliphoridae	sericata	4798
Diptera	Calliphoridae	vomitoria	3443
Diptera	Cecidomyiidae	acarisuga	653
Diptera	Cecidomyiidae	aphidimyza	1307
Diptera	Cecidomyiidae	brassicae	1089
Diptera	Culicidae	aegypti	900
Diptera	Drosophilidae	melanogaster	2676
Diptera	Muscidae	aenescens	3540
Diptera	Muscidae	domestica	3617
Diptera	Syrphidae	corollae	1534
Diptera	Syrphidae	rueppellii	1759
Diptera	Tephritidae	oleae	2923
Diptera	Tipulidae	oleracea	2534
Hemiptera	Aleyrodidae	proletella	740
Hemiptera	Aleyrodidae	vaporariorum	1060
Hemiptera	Cicadellidae	titanus	1596
Hemiptera	Pentatomidae	halys	1534
Hemiptera	Pentatomidae	italicum	1596
Hymenoptera	Andrenidae	vaga	1560
Hymenoptera	Apidae	mellifera carnica	1552
Hymenoptera	Apidae	mellifera iberica	3334
Hymenoptera	Apidae	mellifera ligustica	2935
Hymenoptera	Apidae	mellifera ligustica x	1595
Hymenoptera	Apidae	pascuorum	1831
Hymenoptera	Apidae	terrestris	5021
Hymenoptera	Braconidae	colemani	1047
Hymenoptera	Braconidae	matricariae	1074
Hymenoptera	Megachilidae	rotundata	1721
Hymenoptera	Tenthredinidae	pectinicornis	1517
Hymenoptera	Tenthredinidae	rosae	1516
Hymenoptera	Vespidae	vulgaris	1456
Lepidoptera	Plutellidae	xylostella	1383
Lepidoptera	Tortricidae	botrana	1193
Lepidoptera	Yponomeutidae	padella	1618
Neuroptera	Chrysopidae	carnea	3147
Odonata	Coenagrionidae	elegans	1735

Supplementary Table 1. *The number of observations available from the flight cage recordings and used for method development.*



Supplementary figure 1. *Correlation of Simpson's and Shannon's diversity indices (D and H) respectively- The DBSCAN method is negatively correlated with the traps in both metrics as seen in a) and c). While the HCA methods number are much higher, they show good correlation with the traps b) & d).*

Patent I - Method And Apparatus For Determining An Index Of Insect Biodiversity, An Insect Sensor And A System Of Insect Sensors

US 20230106933A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2023/0106933 A1**

Rydhmer et al. (43) **Pub. Date: Apr. 6, 2023**

(54) **METHOD AND APPARATUS FOR DETERMINING AN INDEX OF INSECT BIODIVERSITY, AN INSECT SENSOR AND A SYSTEM OF INSECT SENSORS**

(71) Applicant: **FaunaPhotonics Agriculture & Environmental A/S**, Copenhagen SV (DK)

(72) Inventors: **Klas Rydhmer**, Malmö (SE); **Thomas Nikolajsen**, Slangerup (DK); **Laurence Still**, Copenhagen S (DK); **Mikkel Brydegaard Sörensen**, Lund (SE); **Flemming Bent Rasmussen**, Frederiksberg (DK); **Alfred Gösta Victor Strand**, Malmö (SE)

(21) Appl. No.: **17/904,422**

(22) PCT Filed: **Feb. 19, 2021**

(86) PCT No.: **PCT/EP2021/054181**

§ 371 (c)(1),
(2) Date: **Aug. 17, 2022**

(30) **Foreign Application Priority Data**

Feb. 19, 2020 (EP) 20158120.4

Publication Classification

(51) **Int. Cl.**

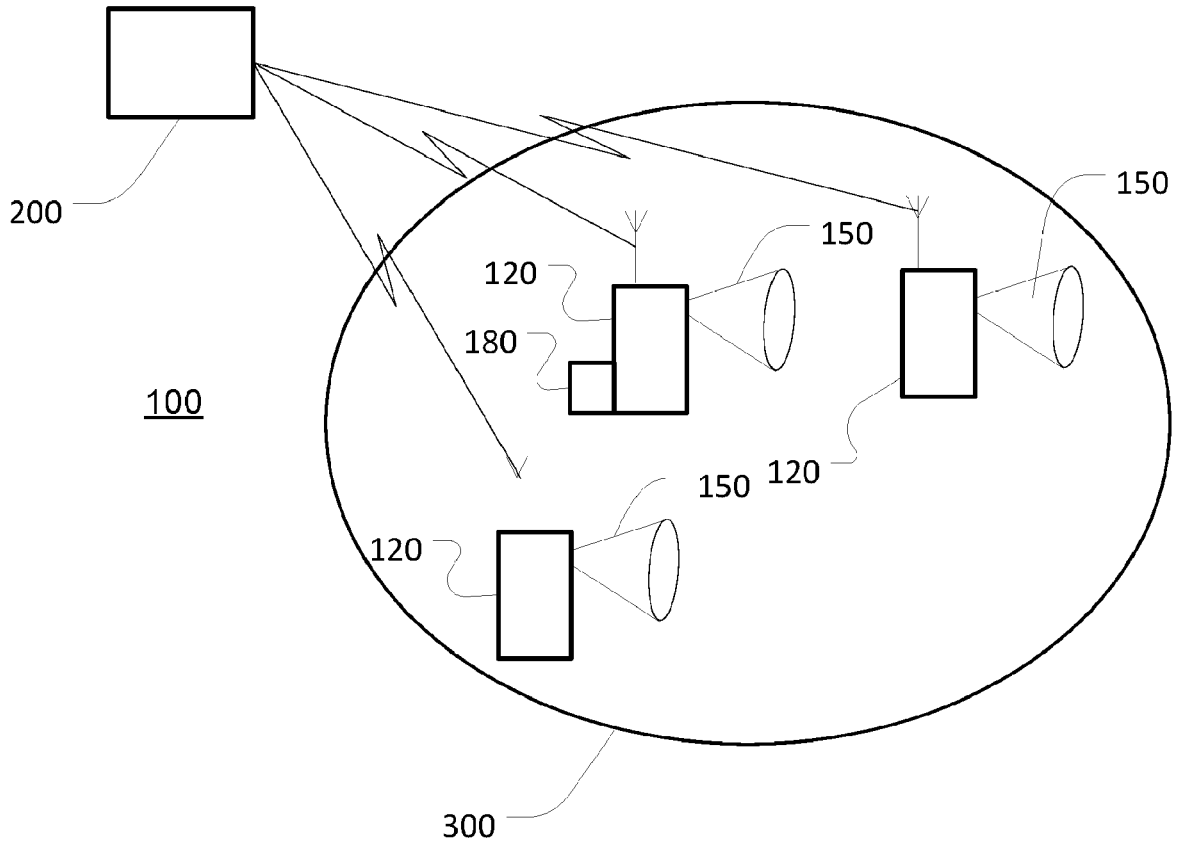
A01K 29/00	(2006.01)
H04N 7/18	(2006.01)
G06V 10/762	(2006.01)
G06V 40/10	(2006.01)

(52) **U.S. Cl.**

CPC ***A01K 29/00*** (2013.01); ***H04N 7/181*** (2013.01); ***G06V 10/762*** (2022.01); ***G06V 40/10*** (2022.01)

(57) **ABSTRACT**

An apparatus for determining an index of insect biodiversity, comprising: a plurality of optical insect sensor devices configured to be individually positioned within a geographic area, each insect sensor device configured to: monitor insect activity within a detection volume extending outside the insect sensor device by detecting light from the detection volume, and to output detector data indicative of one or more optically detected attributes associated with respective detected insect detection events, each insect detection event being indicative of one or more insects being present in the detection volume; a data processing system communicatively coupled to the plurality of optical insect sensor devices and configured to: receive detector data from respective ones of the plurality of optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective detected insect detection events, and to compute, from the received detector data, an index of insect biodiversity indicative of insect biodiversity within the geographic area.



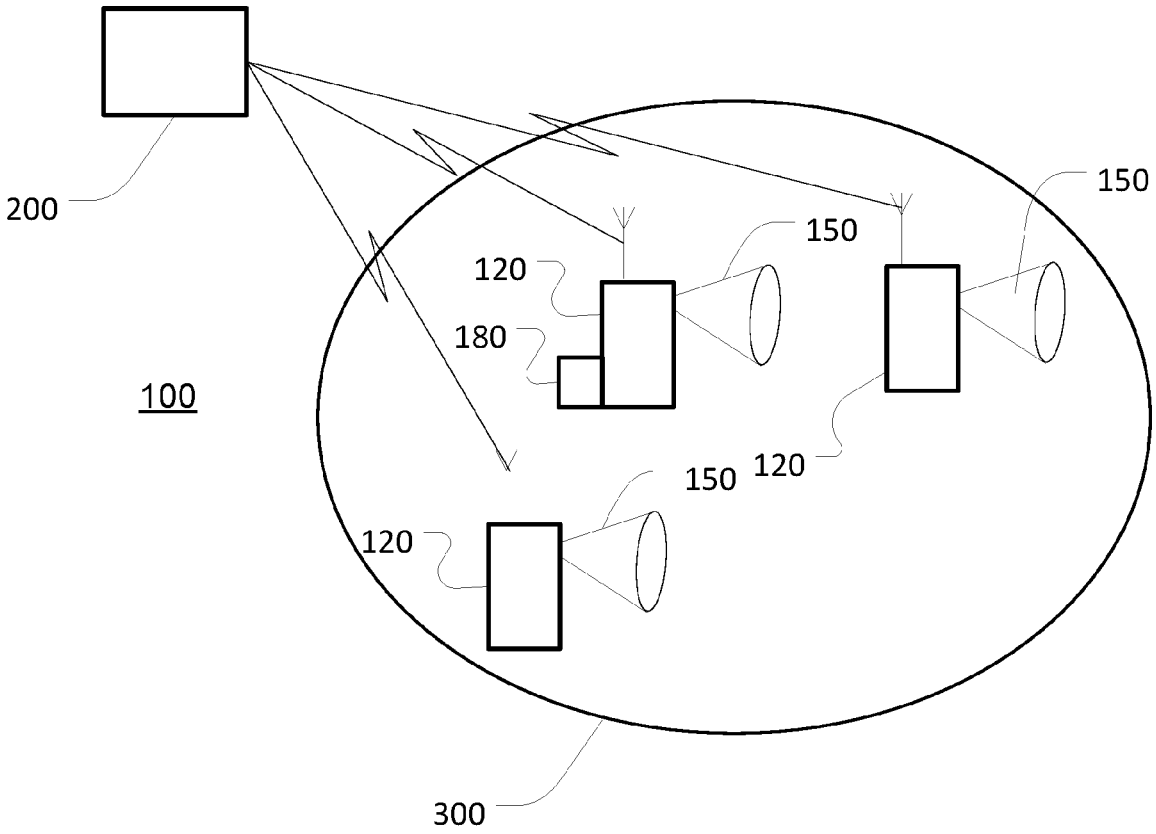


FIG. 1

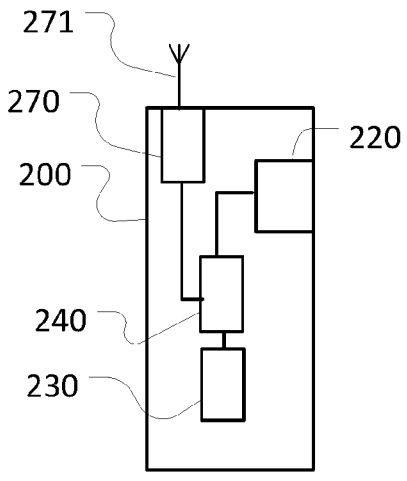


FIG. 2

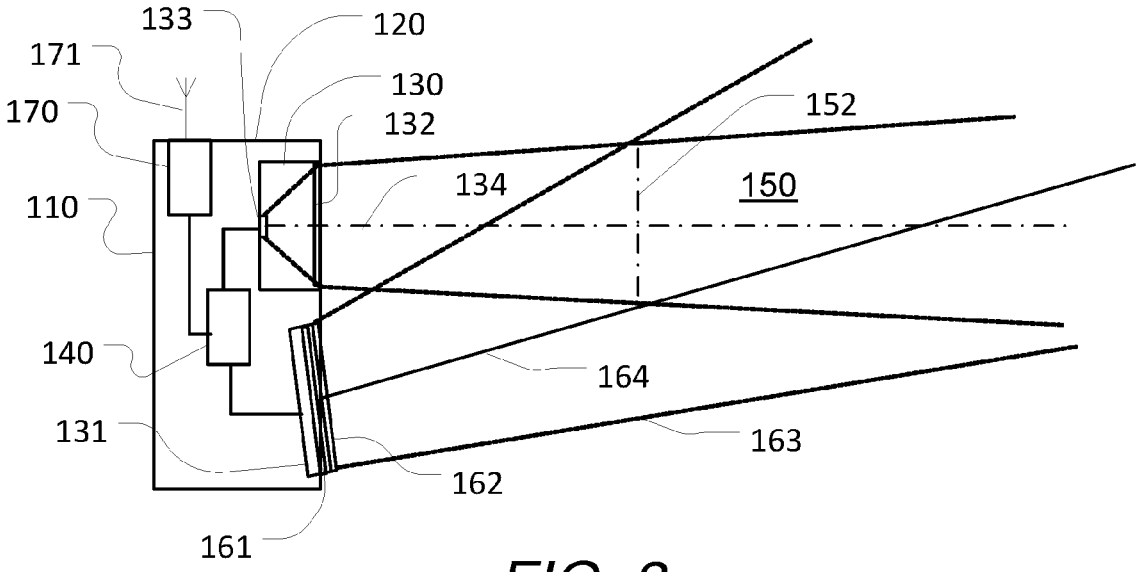

133
120
171
130
170
132
152
134
150
110
140
164
131
163
162
161

FIG. 3

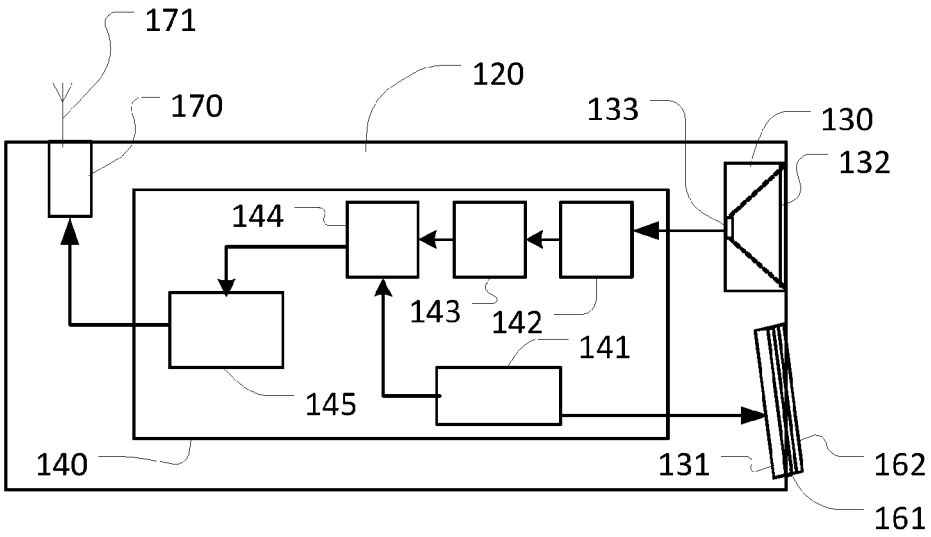

171
120
170
133
130
132
144
143
142
141
145
140
131
162
161

FIG. 4

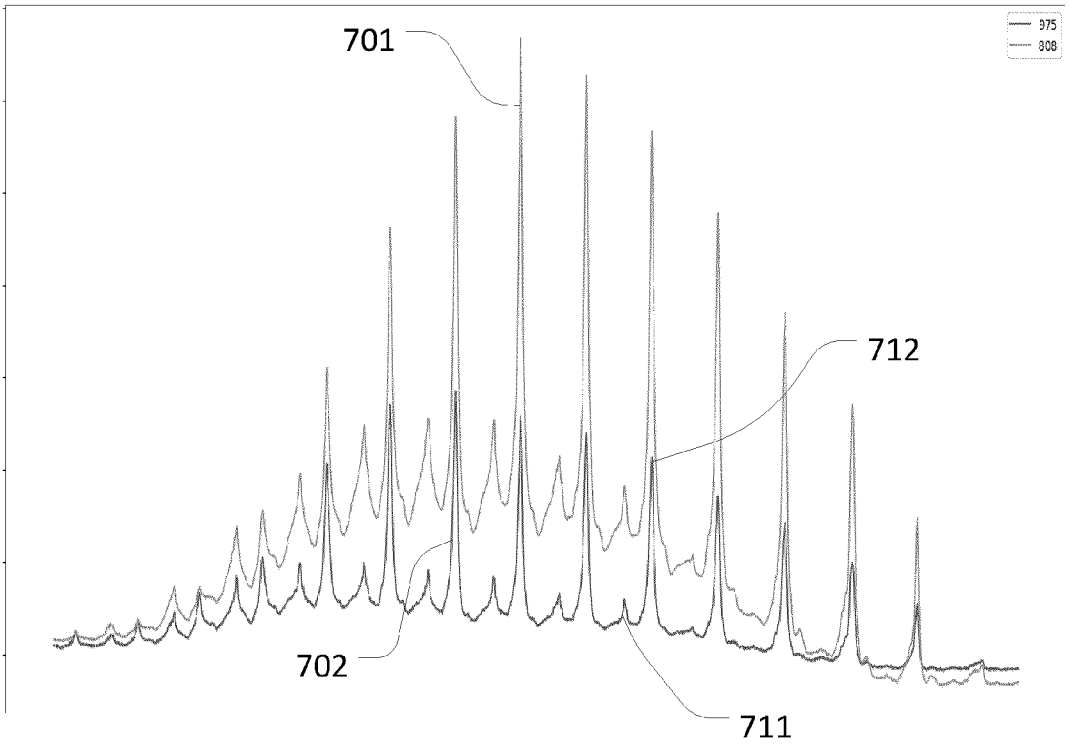


FIG. 5

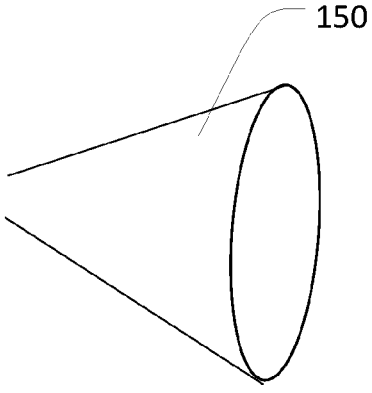


FIG. 6

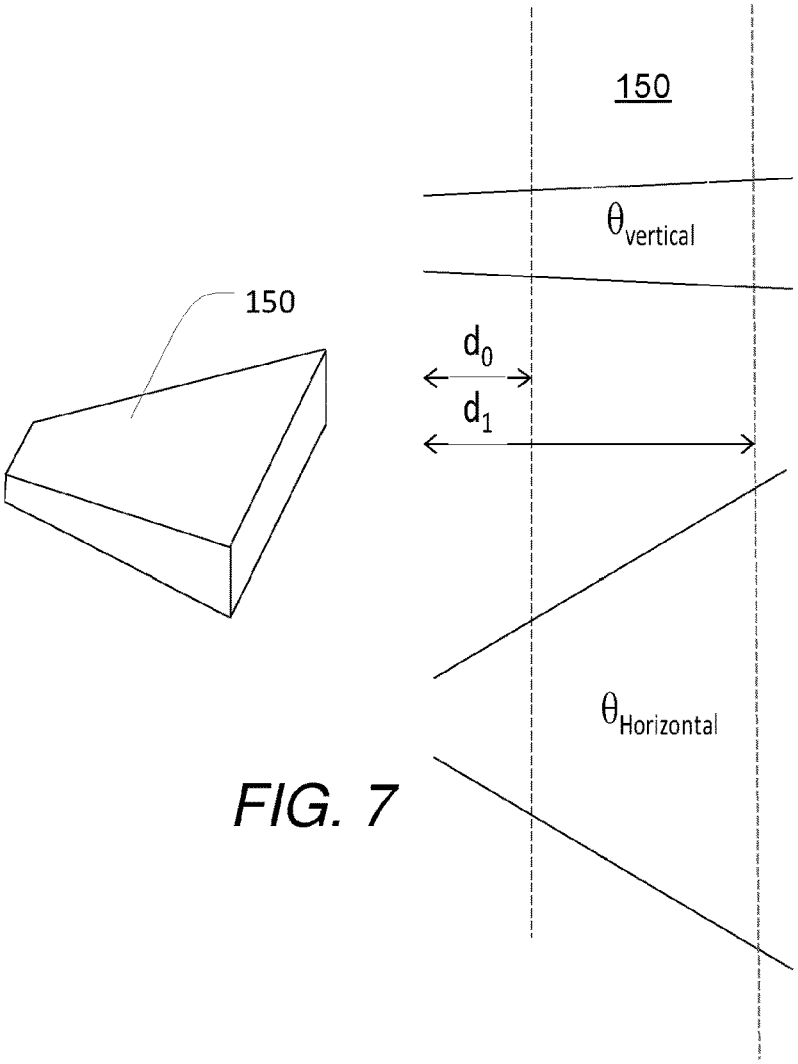

150
150
$\theta_{vertical}$
d_0
d_1
$\theta_{Horizontal}$

FIG. 7

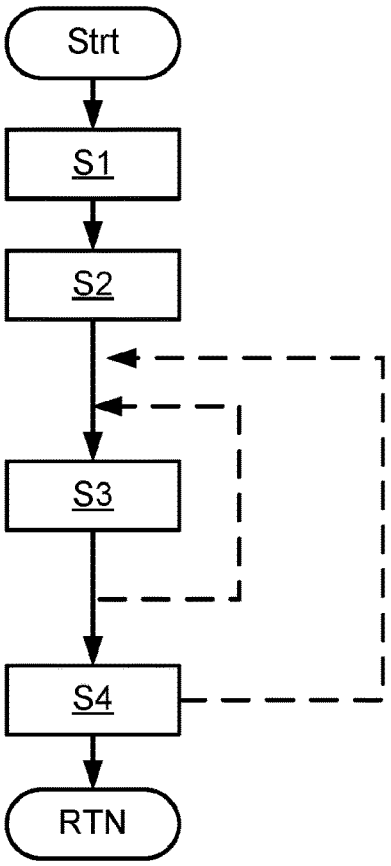

Strt
S1
S2
S3
S4
RTN

FIG. 8

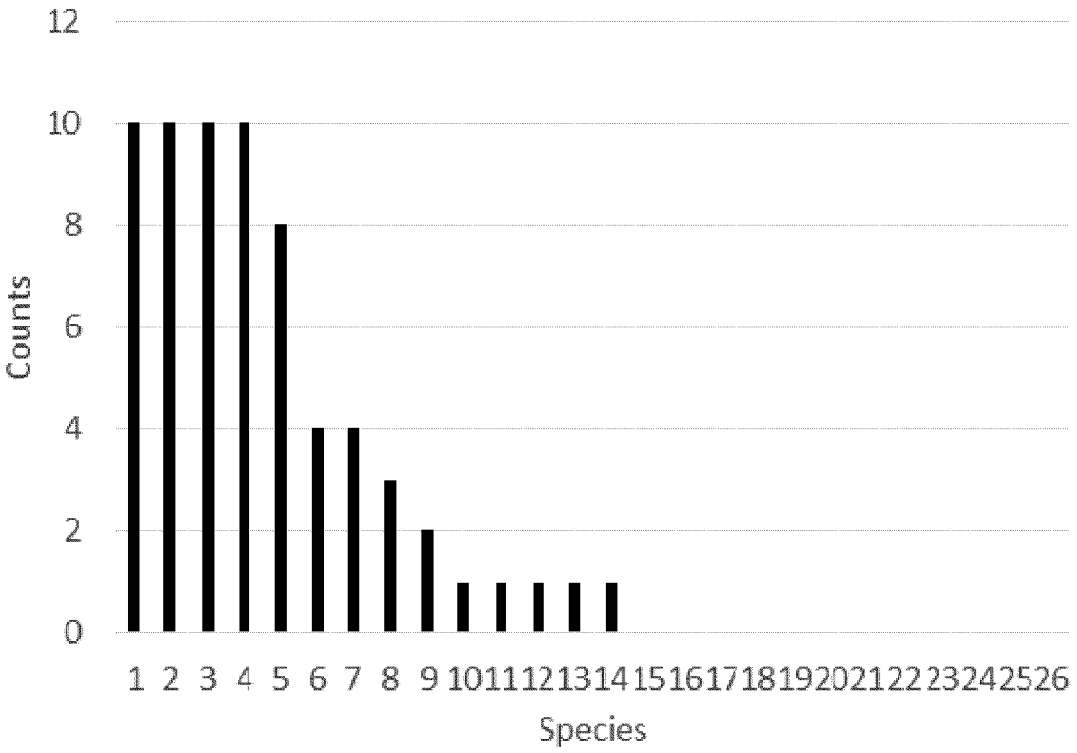

12
10
8
6
4
2
0
Counts
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Species

FIG. 9A

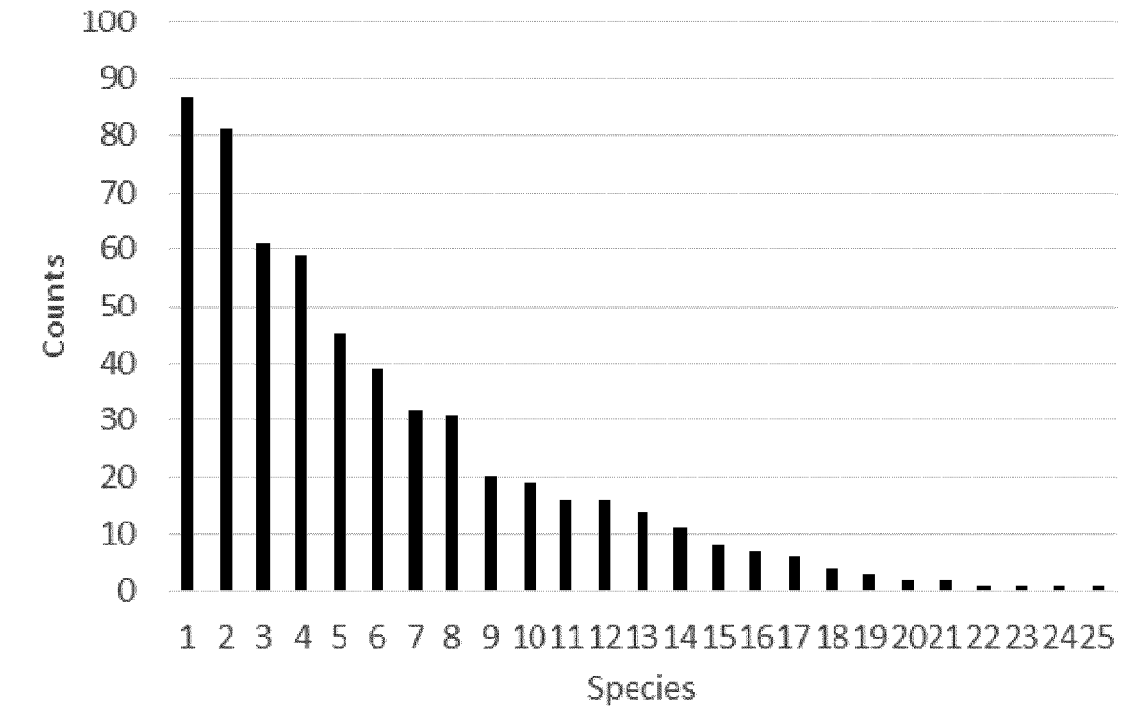

100
90
80
70
60
50
40
30
20
10
0
Counts
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Species

FIG. 9B

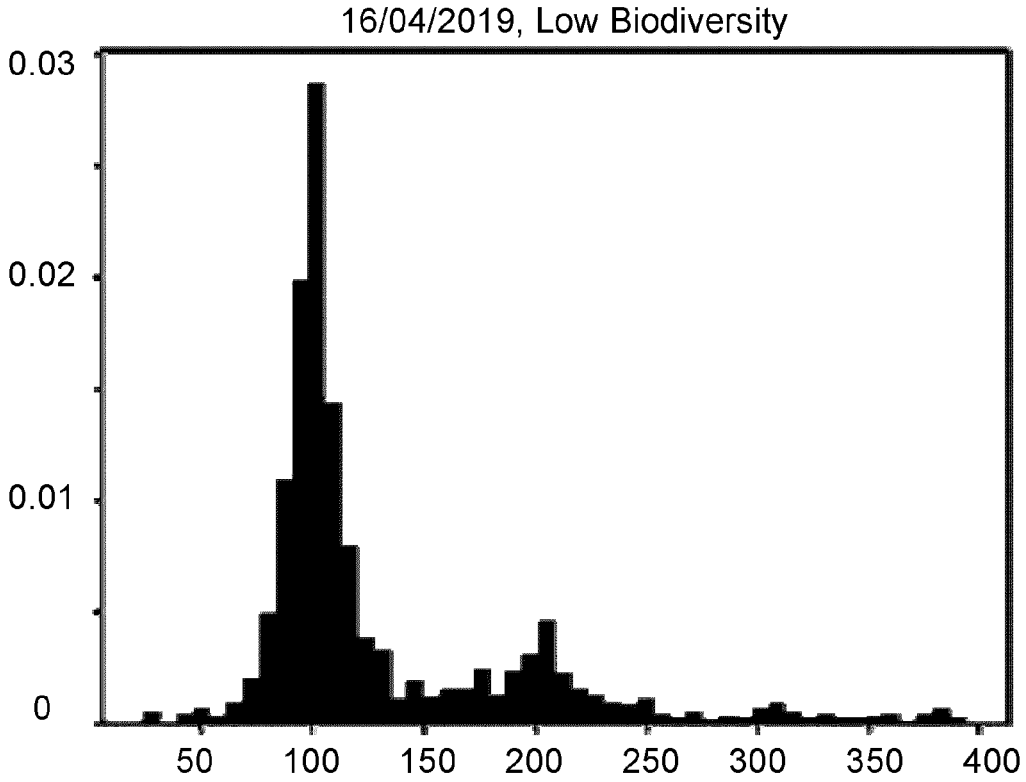


FIG. 10A

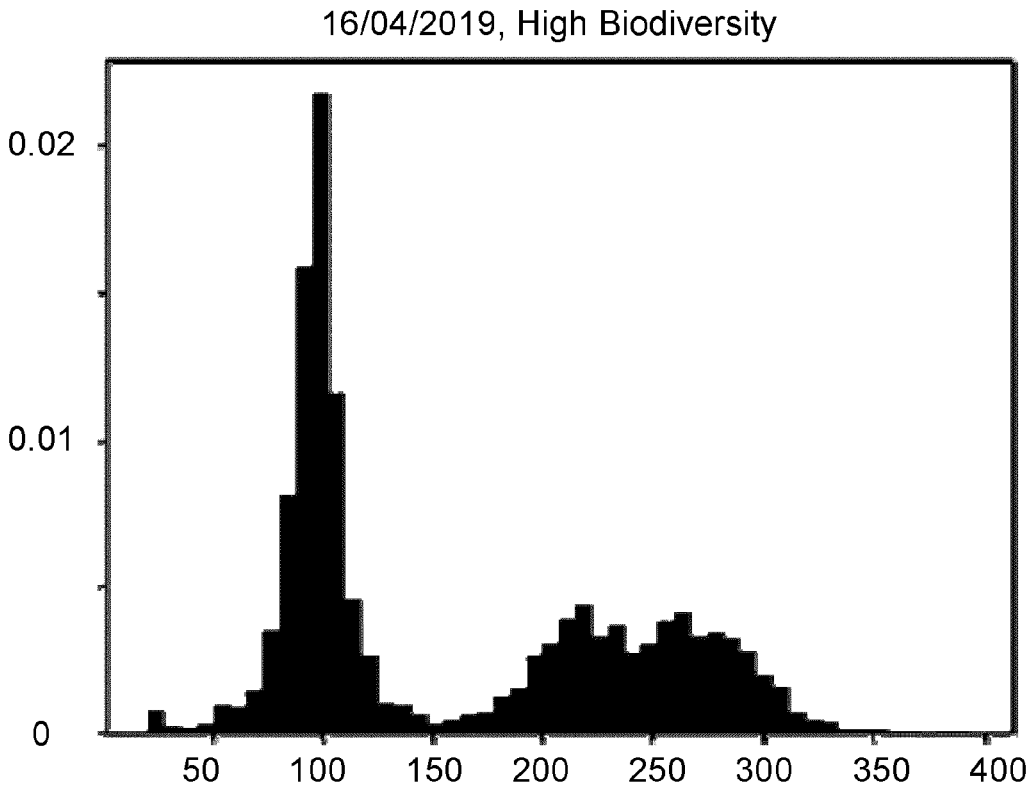


FIG. 10B

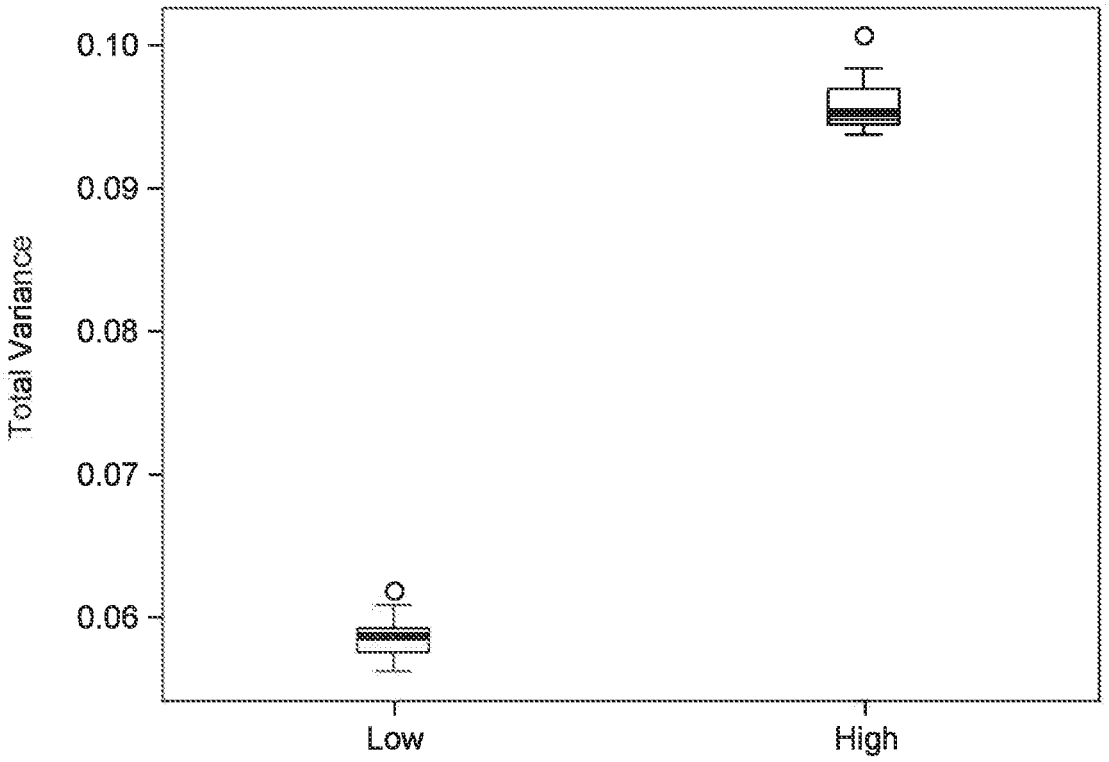


FIG. 11

METHOD AND APPARATUS FOR DETERMINING AN INDEX OF INSECT BIODIVERSITY, AN INSECT SENSOR AND A SYSTEM OF INSECT SENSORS

TECHNICAL FIELD

[0001] The present disclosure relates to a method and to an apparatus for determining an index of insect biodiversity. The present disclosure further relates to an insect sensor and a system of insect sensors.

BACKGROUND

[0002] Insect decline, in particular loss in insect biodiversity is one of the top challenges facing the world today and considered one of the greatest risks of the 21st century.

[0003] A recent OECD report entitled "*Biodiversity: Finance and the Economic and Business Case for Action,* May 2019" states: "Biodiversity loss is one of the greatest risks of the 21st century. It undermines human health and well-being, societal resilience and progress towards the Sustainment and development goals. It places severe costs on our economies and makes addressing other global challenges, such as climate change, much more difficult."

[0004] "The planet is facing its sixth mass extinction, with the current rate of species extinction estimated to be as high as 1000 times the background (pre-human) rate. In addition, widespread and rapid population declines are affecting even common species that are fundamental to ecological processes: since 1970, the world has lost 60% of its global vertebrate population, and more than 40% of insect species are declining rapidly".

[0005] When addressing problems related to insect decline, the current lack of precise and accurate data is a key problem for research, businesses and decision makers. Therefore, there is an urgent need for precise data that may guide and facilitate proper actions and legislation to ensure sustainable environments and use of chemicals, while allowing sufficient food production to support a growing population. Current methods for studying and monitoring biodiversity and insect behaviour rely on laborious methods where insects are collected with nets and manual traps and subsequent analysed via microscopic and genetic analysis.

[0006] It is thus desirable to provide efficient yet precise methods and apparatus for determining an index of insect biodiversity that allows the collection of comprehensive data to support decision-making processes and that provides industry with a tool for balancing production with sustainability. It is further desirable to provide cost-efficient and reliable insect sensors for measuring insect activity.

[0007] The number of insects in a geographic area may vary over time but also across a given area. These temporal and spatial variations make accurate determination of reliable and consistent measures of insect activity and, in particular, insect biodiversity challenging. In particular, insects are often non-uniformly distributed across an area and hot spots of locally high insect concentrations may occur. Moreover, the location of such hot spots may change over time.

[0008] Yet further, traditional methods for quantifying a biodiversity index rely on accurate identification of insect species, thus rendering traditional methods labour intensive, sensitive to variation in expert knowledge and error prone. This in turn reduces the scalability of traditional methods and may reduce their comparative value when calculated for different geographic areas inhabited by different species and/or classified by different human experts. For example, the Institute for European environmental policy summarized the research situation in Denmark in 2017 as follows: "There is a very small number of wild bee experts in Denmark and a significant lack of knowledge on species abundance, distribution and trends" (see E. G. Evelyn Underwood, Gemma Darwin, "Pollinator Initiatives in EU Member States: Success Factors and Gaps," 2017.) Today, research and monitoring are thus hampered by very expensive and labour-intensive practice when studying insects, typically involving sweep-netting, trapping and identification.

[0009] Trapping can be done with various methods depending on the target insect, as each trap type is biased towards different species. The total number of collected insects, N, may be identified by a taxonomist typically using a microscope or DNA sequencing of each individual collected insect.

[0010] From this type of taxonomy, insects may be separated into different groups. The number of groups is typically referred to as the richness of the population, R, while the number of insects in each group is referred to as the abundance, r_i, of insects within a certain group. From these data a biodiversity index can be constructed in various ways such as, the Shannon biodiversity index, H:

$$H = \sum_{i=1}^{R} r_i \ln(r_i)$$

[0011] or the Simpson biodiversity, D

$$D = 1 - \frac{\sum_{i=1}^{R} r_i(r_i - 1)}{N(N-1)}$$

[0012] An important point is that the grouping of insects is not a standardised process but rather depends strongly on the context of a specific study. As examples, species delimitation can be according to order (flies), family (hoover files) and even genus type within a family. Also, studies are performed where the biodiversity is defined according the bio services various groups provide to an ecosystem, e.g. pollination.

[0013] Using the approach outlined above the ability to obtain highly temporally and spatially resolved data is usually very limited and studies are either done intensively over a short period or at low intensity over a long period. For example, a recent insect decline study (see C. A. Hallmann et al., "More than 75 percent decline over 27 years in total flying insect biomass in protected areas," PLoS One, vol. 12, no. 10, 2017) involved malaise traps being emptied on average every 11 days, after which each insect sample was manually dried and weighed. Simply monitoring the flying biomass is therefore a massive task and hard to repeat across different habitats.

[0014] In order to reduce the labour associated with the monitoring of insects, attempts have been made to introduce technical solutions.

[0015] For example, U.S. Pat. No. 9,585,376 describes a system of electronic insect monitoring devices (EIMDs). The EIMDs each comprise a lure for attracting at least one

target insect species, one or more sensors that generate one or more output signals in response to an insect approaching the lure, and an electronic controller configured to determine if the insect approaching the lure belongs to the at least one target insect species using the one or more output signals. In some embodiments, this prior art system may comprise a plurality of EIMDs configured to communicate over a wireless network shared by the plurality of EIMDs. However, lures typically only attract certain species and are therefore unsuitable for monitoring the biodiversity of the total insect population in a given area.

[0016] In a different field, U.S. Pat. No. 5,956,463 describes an automated system for monitoring wildlife auditory data and recording same for subsequent analysis and identification. The system comprises one or more microphones, which may be located at various locations in the field, coupled to a recording apparatus for recording wildlife vocalizations in digital format. The resultant recorded data is pre-processed, segmented, and analyzed by means of a neural network to identify the respective species. In particular classification of the species is used to discriminate wildlife calls and to identify the animal from which a selected call originated. Even though this prior art system minimizes the need for human intervention and subjective interpretation of the recorded sounds, its usability for determining insect biodiversity is limited. Firstly, audio-based instruments are sensitive to audio noise, which may mask sounds generated by insects, thus limiting the detection range of the system and its usability in urban areas, along roads or in the vicinity of other sound sources. Moreover, classification of species based on wildlife calls is not suitable for the majority of insects. Finally, training a neural network classification system that determines species still requires laborious collection and classification of training data.

[0017] The value in monitoring biodiversity is significant for both agriculture and ecology and the inventors are not aware of known tools which can provide a standardised measure of flying insect diversity. It is thus generally desirable to provide a method and apparatus suitable for determining insect biodiversity. In particular it is desirable to provide a method for determining insect biodiversity that is scalable, accurate and efficient. It further remains desirable to provide an apparatus that is low complex, durable, reliable and accurate. Yet further, it is desirable to provide an apparatus that is largely non-intrusive, that allows for a substantially unbiased detection of a large variety of types of insects and that only minimally affects the insect activity, as this allows for a more accurate determination of a biodiversity index.

SUMMARY

[0018] Disclosed herein are embodiments of a method and an apparatus for determining an index of insect biodiversity as well as embodiments of an insect sensor and of a system of insect sensors.

[0019] According to a first aspect, embodiments of an apparatus for determining an index of insect biodiversity comprise:

[0020] a plurality of optical insect sensor devices configured to be individually positioned within a geographic area, each insect sensor device configured to:

[0021] monitor insect activity within a detection volume extending outside the insect sensor device by detecting light from the detection volume, and to

[0022] output detector data indicative of one or more optically detected attributes associated with respective detected insect detection events, each insect detection event being indicative of one or more insects being present in the detection volume;

[0023] a data processing system communicatively coupled to the plurality of optical insect sensor devices and configured to:

[0024] receive detector data from respective ones of the plurality of optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective detected insect detection events, and to

[0025] compute, from the received detector data, an index of insect biodiversity indicative of insect biodiversity within the geographic area.

[0026] As the insect sensor devices optically detect attributes associated with insect detection events in a detection volume outside the insect sensor devices, the apparatus is non-intrusive to the environment in the sense that it does not rely on and, consequently, is not biased by pheromones or other means of attracting, trapping or killing insects. In particular, insects may be detected in their natural environment regardless of their affinity to a certain lure or trap technology, thus reducing the sensitivity of the measurement results to different trapping techniques for different insect species. To this end, the detection volume is preferably an enclosure-free void/space allowing unrestricted movement of living airborne insects into and out of the void/space.

[0027] According to a second aspect, embodiments of an apparatus for determining an index of insect biodiversity comprise:

[0028] one or more optical insect sensor devices configured to be individually positioned within a geographic area, each of the one or more insect sensor devices configured to:

[0029] monitor insect activity within a detection volume to detect one or more insect detection events, each insect detection event being indicative of one or more insects being present in the detection volume; and to

[0030] output detector data indicative of one or more optically detected attributes associated with respective detected insect detection events,

[0031] a data processing system communicatively coupled to the one or more optical insect sensor devices and configured to:

[0032] receive detector data from the one or more optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective detected insect detection events, wherein the detected insect detection events are taxonomically unclassified, and to

[0033] compute an index of insect biodiversity directly from the optically detected attributes associated with the taxonomically unclassified detection events.

[0034] Accordingly, the data processing system is configured to compute the biodiversity index directly from the optically detected attributes, i.e. without an intermediate taxonomical classification of the individual detected insects, i.e. without mapping the detected insects to known groups of insects, e.g. to known families or species. In other words, the step of insect taxonomic classification is avoided and the

biodiversity index is calculated directly from the collected data samples instead. Embodiments of the apparatus may thus automatically collect and process detector data in a standardized way, thereby increasing comparability of the results across different geographic areas. Moreover, the process reduces the influence of misclassifications, e.g. due to human errors, on the resulting index of insect biodiversity. Yet further, the apparatus is capable of computing an index of insect biodiversity even in cases where the insect species of the population are not known or where the optically detectable attributes have not previously been classified to known groups of insects.

[0035] Generally, embodiments of the apparatus according to one or both of the above aspects allow a more uniform detection of a large variety of insects, regardless of the specific insect species. Hence, embodiments of the apparatus provide a computation of a more accurate measure of insect biodiversity. As the apparatus automatically computes the index of insect biodiversity from the acquired detector data, the apparatus is less sensitive to researcher bias or variation in classification technique.

[0036] Embodiments of the apparatus described herein can provide a standardised measure of insect biodiversity to allow comparison of biodiversity over time and between different geographic locations. Embodiments of the apparatus may be operated in a fully automated manner, thus facilitating assessment of insect biodiversity in real time, day by day, hour by hour and even minute by minute.

[0037] The computation of the index of insect biodiversity directly from the detector data may be done in a number of ways.

[0038] In particular, computing the index of insect biodiversity may comprise computing the index of insect biodiversity from the optically detected attributes, which are associated with respective ones of a set of detected insect detection events. In some embodiments, each element of the set of detected insect detection events may be associated with a plurality of attributes of different types, e.g. a wing beat frequency and a melanisation ratio and/or other attributes associated with the insect detection event. For example, each insect detection event may have associated with it an n-tuple. Each n-tuple represents n attributes associated with the corresponding insect detection event. The process may use different numbers of attributes, i.e. n may be an integer larger or equal than 1, e.g. n may be equal to 1, 2, 3, 4 or larger. For example, in one embodiment n=2 and each n-tupel includes a wing beat frequency and a body-wing ratio. The set of detected insect detection events from which the index of biodiversity is calculated includes a plurality of insect detection events, such as at least 50 insect detection events, e.g. at least 100 insect detection events, such as at least 500 insect detection events.

[0039] In some embodiments, computing the index of insect biodiversity comprises computing the index of insect biodiversity as a measure of variability of the optically detected attributes associated with respective ones of the set of detected insect detection events. In particular, the data processing system may compute the measure of variability by performing a statistical analysis of the optically detected attributes associated with the set of detected insect detection events, in particular a multi-variate analysis. For example, the data processing system may compute one or more measures that represent the diversity of the set of n-tuples. Example measures of variability include but are not limited to a covariance trace (total variance), a covariance entropy, a covariance determinant and a summed standard deviation, or a combination thereof.

[0040] In some embodiments, computing the measure of variability includes performing a clustering of the detected insect detection events according to at least the optically detected attributes associated with the respective detected insect detection events. The clustering may result in a set of clusters, each cluster including one or more of the detected insect detection events. Example clustering approaches include DBSCAN and Gaussian Mixture Models. Useful outputs of clustering-based approaches include but are not limited to the number of clusters, required parameters to achieve a given number of clusters or features of the Akaike or Bayesian information criterion.

[0041] There are a number of ways the process may compute an index of insect biodiversity from an output of an unsupervised clustering process. For example, the index of insect biodiversity may be computed from the number of resulting clusters, by computing a Shannon or Simpson index from the resulting clusters, by applying a Gaussian mixture method, by applying a Gaussian mixture AIC gradient approach, etc., or a combination thereof.

[0042] Using for example the DBSCAN clustering process, the number of clusters found is one of the output parameters of the clustering process. This number of clusters can be used as a biodiversity index as a number reminiscent of species richness (i.e. the number of different species present in an area).

[0043] Other clustering models, such as Gaussian Mixture Models, require the number of clusters to be pre-defined before training, and therefore do not a priori provide a richness-like measure. One approach here is to pre-define a range of possible clusters (e.g. between 1 and 100 with a step of 1) and then fit the model for each cluster number. The goodness-of-fit of these models can be described by an external function for example the Bayesian Information Criterion or the Akaike Information Criterion. Thus, a similar 'best' number of clusters can be determined by identifying the minimum point in the generated distribution of goodness-of-fit values.

[0044] The Shannon and Simpson indices are methods used to quantify the diversity of a population which can be divided into discrete groups. In conventional ecology these groups are generally aligned with taxonomic levels, such as species, genus or family. However, in an unsupervised approach to biodiversity quantification, the optimum number of clusters—e.g. determined as discussed above—may be used. In this case, the input parameters to the Shannon and Simpson indices would be the number of clusters, and the number of individuals sorted into each cluster.

[0045] The BIC and AIC are not infallible methods, and sometimes do not produce a single, or reliable, minimum value. However, in the case of the BIC the distribution at some point always crosses the x axis at a value which changes with the number of ideal clusters. The more clusters present in the data, the higher the distribution crosses the x axis. This approach always results in a single value (the x axis is only crossed once) and the value is much higher than the predicted number of clusters which allows for some additional nuance. This approach has been found by the inventors to produce values which correlate well with the ground truth data with a correlation coefficient of R2=0.90.

[0046] The AIC gradient approach is an attempt to improve on the Akaike Information Criterion approach to assessing the optimal number of clusters. The AIC function, when applied to a dataset descriptive of insect detection events, often takes the form of an oblique elbow, where the criterion drops sharply as the number of clusters increases, and then at some inflection point the gradient sharply decreases and the criterion continues to decline. This means that there is no defined minimum point. The gradient of the AIC appears to correlate with the point that the elbow inflection lay, so this approach quantifies the gradient of the initial linear decline.

[0047] Other examples of methods for computing a measure of variability of the optically detected attributes associated with respective insect detection events include a statistical description of raw events using auto-encoding or similar techniques and dimensional reduction of the optically detected attributes.

[0048] In some embodiments, the computation of the index of insect biodiversity is based on a mathematical model that directly maps the optically detected attributes associated with respective detected insect detection events to an index of insect biodiversity. In particular, the mathematical model may be a machine learning model such as a supervised machine learning model. For example, when applying a supervised machine learning model, the data processing system is configured to implement a trained machine learning model, in particular a model trained by supervised learning. The supervised learning may involve applying a training set to train a machine learning model to predict the biodiversity of an input either as a number such as an existing or novel index, or into one or more categories. The training data may comprise detector data from one or more insect sensor devices and known biodiversity measurements from either field trials or simulated data. Examples of machine learning models include a convolutional or fully-connected neural network, a decision tree, or the like.

[0049] Generally, the index of insect biodiversity refers to a suitable numerical measure of insect biodiversity, in particular airborne insects such as flying or jumping insects. For the purpose of the present description an index of insect biodiversity will also be referred to as insect biodiversity index or merely biodiversity index. The index of insect biodiversity mat be represented as a number, e.g. a number between 0 and 1, or in another suitable way, e.g. as a categorization into biodiversity classes, e.g. "low', "medium", "high", etc. In some embodiments, the computation of the index of insect biodiversity may comprise correlating the optically detected attributes associated with a set of detected insect detection events to a known biodiversity metrics such as, but not limited to, the Simpsons or the Shannon biodiversity index.

[0050] Each insect sensor device may be mounted at a stationary detection site or non-stationary, e.g. mounted on a vehicle. The vehicle may be a ground vehicle, i.e. a vehicle that operates while in contact with the ground surface. A ground vehicle may e.g. drive on wheels or the like. For example, the ground vehicle may be a tractor or other farming vehicle. Other examples of vehicles include aerial vehicles such as an airplane, a helicopter or the like. The vehicle may be a manned vehicle or an unmanned vehicle.

[0051] Each insect sensor device may comprise an illumination module configured to illuminate the detection volume, in particular the entire detection volume. Each insect sensor device may comprise a detector module comprising one or more detectors configured to detect light from the detection volume, in particular from the entire detection volume. The detector module may thus output a sensor signal indicative of the detected light, e.g. indicative of a detected light intensity as a function of time.

[0052] In some embodiments, each insect sensor device comprises an illumination module and a detector module.

[0053] The apparatus may comprise one or more processing units configured to receive a sensor signal from a detector module of at least one of the insect sensor modules and to process the received sensor signal so as to detect one or more insect detection events and to extract one or more optically detectable attributes associated with the detected insect detection events. The processing unit may be implemented as a local processing unit, integrated into an insect sensor device and configured to process sensor signals of the detector module of said insect sensor device into which the processing unit is integrated. In other embodiments, some or all of the processing steps are performed by a processing unit external to the insect sensor device, i.e. the processing unit may be implemented in a device external to the insect sensor device or it may be distributed between a local processing unit of the insect sensor device and a remote processing unit, separate from the insect sensor device

[0054] In some embodiments, the illumination module comprises a light source that is configured to emit incoherent light. Suitable light sources include light-emitting diodes (LEDs) and halogen lamps, as these are able to simultaneously illuminate a large detection volume with sufficient light intensity. Further, incoherent light sources are useful to provide a homogeneous, speckle free, illumination of the detection volume, in particular a simultaneous illumination of a large detection volume without the need for any scanning operation. This reduces the complexity of the optical system and allows reliable detection of wing beat frequencies and/or trajectories even of fast-moving insects.

[0055] Nevertheless, other light sources, including coherent light sources, such as lasers, may be used instead. In some embodiments, the light source is configured to output light continuously while, in other embodiments, the light is turned on and off intermittently, e.g. pulsed.

[0056] In some embodiments, the illumination module is configured to emit light with varying intensity, in particular pulsed or otherwise modulated at one or more modulation frequencies.

[0057] Some embodiments of the apparatus include one or more insect sensor devices disclosed in the following, in particular one or more embodiments of the insect sensor device according to the third aspect described below and/or a system of insect sensor devices according to the fourth aspect described below.

[0058] According to a third aspect, disclosed herein are embodiments of an optical insect sensor device for detecting insects in a detection volume, the insect sensor device comprising:

[0059] an illumination module configured to illuminate the detection volume with illumination light comprising light at a first wavelength band modulated at a first modulation frequency and light at a second wavelength band;

[0060] a detector module comprising a detector configured to detect light from the detection volume;

[0061] a processing unit configured to receive sensor signals from the detector module and configured to filter the received sensor signal to extract a first sensor signal modulated at the first modulation frequency and, based on at least the first sensor signal, to detect at least one insect in the detection volume and to determine at least one optically detectable attribute of the detected insect, such as a melanisation ratio and/or a direction of movement.

[0062] Embodiments of the insect sensor device provide accurate measurements while maintaining a low optical complexity of the insect sensor device. Embodiments of the insect sensor device allow accurate measurements of spectral reflectivity at one, two or more wavelengths, largely unaffected by background illumination such as sunlight, while still allowing a high temporal fill factor in one, two or more channels. Accordingly, embodiments of the insect sensor device are particularly useful for measuring insect biodiversity or for performing other insect measurement tasks, as multiple relatively low-cost sensors can be deployed and operated in a large variety of environmental conditions, thus allowing data to be collected from a large total detection volume.

[0063] The insect sensor device may be mounted at a stationary detection site or non-stationary, e.g. mounted on a vehicle, e.g. as described above.

[0064] The illumination module may be configured to illuminate the detection volume with illumination light and the detector module may be configured to detect a backscattered portion of the illumination light, the backscattered portion being backscattered by insects moving about the detection volume. The inventors have found that a reliable detection and/or identification of insects can be performed by detecting and analyzing light, in particular backscattered light, from illuminated insects. In particular, the detector module may be configured to record a temporal profile of the reflected/backscattered light, as the temporal profile of the reflected/backscattered light is a fingerprint of the insect which can be used to distinguish between different types of insects.

[0065] The detection volume is a 3D volume from which the insect sensor device obtains sensor input suitable for the detection of insects. The detection volume may thus completely or partly be defined by the field of view and depth of field of the detector module. In embodiments where the detection volume is illuminated by an illumination module, the detection volume may be defined as an overlap of the volume illuminated by the illumination module and by a volume defined by the field of view and depth of field of the detector module.

[0066] The detection volume may have a predetermined shape, size and position relative to the illumination module and/or relative to the detector module, e.g. relative to an aperture and/or an optical axis of the detector module. In particular, the detection volume may, during the entire detection process, be stationary relative to the detector module and to the illumination module. Accordingly the detector module may comprise one or more lenses that define an optical axis of the detector module and and/or that define a focal length. The focal length may be fixed during the entire detection process. Moreover, the optical axis may be fixed, during the entire detection process, e.g. relative to the illumination module and/or relative to a housing of the insect sensor device. However, it will be appreciated that the insect sensor device may allow the size, shape and/or relative position of the detection volume to be pre-configured and adapted to a specific measurement environment, e.g. by changing a relative position and/or orientation of the illumination module and the detector module. The detector module may further comprise an aperture.

[0067] The detection volume may have a variety of shapes and sizes, such as box-shaped, cylindrical, ball-shaped, cone-shaped, pyramidal, frusto-conical, frusto-pyramidal, etc. In some embodiments, the detection volume has a size of at least 10 l, such as at least 20 l, such as at least 0.2 m^3, such as at least 0.5 m^3, such as at least 1 m^3, such as at least 2 m^3, such as at least 3 m^3. Even when each individual insect sensor device has a relatively small detection volume, e.g. less than 100 l, a system of several individual insect sensor devices may be deployed so as to provide a larger total detection volume in which insect activity can be monitored by the system of such insect sensor devices.

[0068] In some embodiments, the detection volume has a size of less than 20 m^3, such as less than 10 m^3, such as at less than 5 m^3, such as less than 1 m^3, such as less than 100 l, thereby facilitating uniform illumination at high brightness of the entire detection volume while allowing for reliable detection of e.g. trajectories and/or wing beat frequencies of insects.

[0069] In some embodiments, the detection volume has an aspect ratio, e.g. defined as a ratio of a largest edge to a smallest edge of a minimum bounding box of the detection volume. In some embodiments, the aspect ratio is no larger than 10:1, such as no larger than 5:1, such as no larger than 3:1, such as no larger than 2:1. For example, the aspect ratio may be between 1:1 and 10:1, such as between 1:1 and 5:1, such as between 1:1 and 3:1, such as between 2:1 and 3:1. A low aspect ratio of the detection volume allows moving insects to be tracked over a relative long period of time, regardless of the direction of travel of the insects, thus allowing more accurate detection of different insects, e.g. flying or jumping insects, insects moving at different speeds, etc. Moreover, a relatively long observation time also increases the accuracy of the determined optically detectable attributes such as wing beat frequency, etc. The minimum bounding box may have a vertical and two horizontal edges. The vertical edge may be the smallest edge of the minimum bounding box. For example, a ratio between each of the horizontal edges and the vertical edge may be between 2:1 and 10:1, such as between 2:1 and 5:1, such as between 2:1 and 3:1.

[0070] Each insect sensor device may be configured to be deployed in a geographic area such that the detection volume may be elevated above the ground surface by a minimum vertical offset. To this end, the insect sensor device may include or otherwise be configured to be mounted on a stationary or movable support structure. In some embodiments, the insect sensor device and/or the support structure is/are configured such that the detection volume extends from a top of a vegetation canopy upwards. Accordingly, interference of the vegetation with the insect sensor device, e.g. by blocking the light path, is thus avoided or at least reduced. To this end, the minimum vertical offset may be predetermined, e.g. configurable prior to use. To this end, the support structure may be adjustable so as to adjust a mounting height of the insect sensor device, so as to adapt the minimum vertical offset to the vegetation in vicinity of the insect sensor device. For example, the insect sensor device may be mounted such that the vertical offset of the

insect sensor device above the ground surface is adjustable and/or such that the orientation of the insect sensor device relative to the ground surface is adjustable. The size of the vertical offset may depend on the height of the vegetation growing in the area of land where the insect sensor device is deployed. The vertical offset may be chosen to be larger than a height of the vegetation, e.g. larger than a maximum height of population of plants making up the vegetation in the area of land where the insect sensor device is deployed, or larger than a median height of population of plants in the area of land where the insect sensor device is deployed. For example, the minimum vertical offset may be chosen between 10 cm and 5 m, such as between 20 cm and 3 m, such as between 20 cm and 2 m, such as between 50 cm and 2 m. In some embodiments, when a system or apparatus comprises a plurality of insect sensor devices, the vertical offset may be chosen to be substantially the same for all insect sensor devices of the system or apparatus.

[0071] In some embodiments, the detection volume is located in a proximity of the insect sensor device. In particular, the detection volume may extend between a proximal end and a distal end of the detection volume, relative to the insect sensor device, e.g. relative to an aperture or other optical input port of the detector module. In some embodiments, the distal end may be no more than 5 m from the insect sensor device, such as no more than 4 m, such as no more than 3 m. The proximal end may be separated from the insect sensor device, e.g. from an aperture or other optical input port of the detector module, by 1 cm or more, such as by 10 cm or more, such as by 20 cm or more, such as by 30 cm or more.

[0072] Embodiments of the insect sensor device described herein are particularly suitable for detecting airborne insects, such as flying or jumping insects. Embodiments of the insect sensor device described herein allow for detection of insects moving within the detection volume during sufficiently long observation times so as to reliably identify and distinguish different optically detectable attributes, e.g. wing beat frequencies and/or a trajectories and/or body wing ratios and/or melanisation ratios. Such techniques have been found to allow reliable computation of an index of insect biodiversity when individual insects remain in the detection volume sufficiently long.

[0073] Using a wide volume (in particular a volume having a large cross-sectional area when viewed along an optical axis of the detector module) close to the insect sensor device rather than e.g. a narrow beam extending far away from the insect sensor device provides a number of advantages:

[0074] Optical alignment is less sensitive, allowing the insect sensor device to operate for extended periods of time without continuous calibration.

[0075] Lower light intensity is required, which makes the insect sensor device eye safe and therefore capable of unsupervised operation in the field.

[0076] As the received intensity from an insect decreases with distance, a better size estimation can be done based on the amplitude of the received signal.

[0077] Only a low spatial resolution is needed (fewer "pixels") to adequately focus the illuminated volume onto an image sensor. While long-range instruments depend on pulsed light and use very expensive and sensitive photo multiplier tubes to collect the light, or for the geometrical Scheimpflug configurations CMOS line arrays with thousands of pixels, embodiments of the insect sensor described herein are capable of collecting the light on few, e.g. four, photodiodes where each diode sub-samples the beam. This in turn allows use of very high sampling frequencies. The high sampling frequency (MHz range) allows the use of lock-in amplification which enables the ability to record two wavelengths on one diode and/or to suppress background illumination. This reduces the optical complexity of the insect sensor device and allows accurate measurements of spectral reflectivity at one, two or more wavelengths, unaffected by background illumination such as sunlight, while still allowing a high temporal fill factor in two (or more) channels.

[0078] Insect movement in the volume may be monitored in multiple, e.g. 4, sensors with non-overlapping field of view. This means that it is possible to monitor flight direction and speed.

[0079] A wider beam yields longer transit times and insect events which improves data quality.

[0080] In some embodiments, the illumination module comprises a light source that is configured to emit coherent or incoherent visible light and/or infrared and/or near-infrared light and/or light in one or more other wavelength bands. Infrared and/or near-infrared light (such as light in the wavelength band between 700 nm and 1500 nm, such as between 700 nm and 1000 nm) is not detectable by many insects, and thus does not influence the insect's behavior.

[0081] In some embodiments, the illumination module is configured to selectively illuminate the detection volume with light of two or more wavelength bands, in particular two or more mutually spaced-apart wavelength bands. In some embodiments, the illumination module is configured to emit an illumination beam that includes a homogenous mix of light consisting of two or more wavelength bands (e.g. at 808 and 970 nm, respectively). To this end, the illumination module may include a first light source, e.g. comprising one or more LEDs, configured to selectively emit light of a first wavelength band. The illumination module may further include a second light source, e.g. comprising one or more LEDs, configured to selectively emit light of a second wavelength band, which may be spaced-apart from the first wavelength band. The illumination module may further include one or more further light sources, e.g. comprising one or more LEDs, configured to selectively emit light of one or more further second wavelength bands, which may be spaced-apart from the first wavelength band and/or second wavelength band. The detector module may be configured to selectively detect the selected wavelength bands. In one embodiment, the illumination module is configured to emit light at a first wavelength band at 808 nm+/−25 nm and light at a second wavelength band at 970 nm +/−25 nm. Such a multi-spectral illumination system facilitates color detection of moving insects.

[0082] Using more than one wavelength introduces the possibility to measure "colour", or melanin content of targets. Melanin absorption decreases with increasing wavelengths. By using two, or more, channels at well-separated wavelengths, and comparing the ratio of the received intensity at these two wavelengths, the melanin content can be estimated. For example, a possible measure indicative of the melanisation ratio may be expressed as the ratio $I_{\lambda 1}/(I_{\lambda 1}+I_{\lambda 2})$ or another measure of the relative detected intensity (denoted $I_{\lambda 1}$ and $I_{\lambda 2}$, respectively) at two wavelength bands

around wavelengths λ_1 and λ_2, respectively. In one embodiment λ_1=808 nm and λ_2=970 nm.

[0083] In some embodiments, the apparatus further separates the body and wing contribution of the recorded signal. Accordingly, the apparatus may determine both body and wing melanisation in the insect. It also allows the apparatus to more accurately estimate other features such as wingbeat frequency, since the apparatus can treat the signal received in each wavelength independently and get two separate measurements on the wingbeat frequency. If they for some reason do not agree, that insect detection event can e.g. be discarded as noise.

[0084] In some embodiments, the illumination module is configured to illuminate the detection volume with illumination light at the first wavelength band modulated at the first modulation frequency and light at a second wavelength band, different from the first wavelength band, modulated at a second modulation frequency, different from the first modulation frequency. In one embodiment the light at the first wavelength band (e.g. at 808 nm) is modulated at about 80 kHz and the light at the second wavelength band (e.g. at 980 nm) is modulated at about 120 kHz. The detector module may thus be configured to detect light signals from the detection volume and to selectively filter the detected light signals with the first and second frequencies, respectively. In particular, the processing unit of the insect sensor device may be configured to filter the received sensor signal to extract the first sensor signal modulated at the first modulation frequency and a second sensor signal modulated at the second modulation frequency and, based on the first and second sensor signals, to detect at least one insect in the detection volume and to determine at least one optically detectable attribute of the detected insect, such as a melanisation ratio and/or a direction of movement.

[0085] Accordingly, the detector module may selectively detect the respective wavelength bands with a single detector and efficiently suppress background light, such as daylight or light from light sources other than the illumination module. It will be appreciated that alternative embodiments will include more than two wavelength bands, e.g. 3 or 4 or 5 or more wavelength bands, and the more than two wavelength bands may be modulated at respective modulation frequencies. The processing unit may thus be configured to filter the received sensor signal to extract a corresponding plurality of two or more than two sensor signals modulated at the respective modulation frequencies and, based on the more than two sensor signals, to detect at least one insect in the detection volume and to determine at least one optically detectable attribute of the detected insect.

[0086] In one embodiment, the selective filtering of the one or more modulation bands can be done effciently using phase sensitivity lock-in detection which further eliminates noise from other light sources.

[0087] A convenient illumination of a relatively large detection volume, in particular a simultaneous illumination of the entire detection volume, with a compact illumination module, may e.g. be provided when the illumination module is configured to emit a diverging beam of light, in particular a beam of light having a divergence angle in at least one direction of between 2° and 45°, such as between 10° and 30°, or between 35° and 45°, measured as a full angle between rays originating from the light source and intersecting opposite ends of a beam diameter.

[0088] The illumination module may e.g. include one or more optical elements, such as one or more reflectors and/or one or more lenses that direct the light from the light source as a beam of light, such as a diverging beam of light, of a suitable cross-sectional shape towards the detection volume. For example, the beam of light may have a rectangular or round, e.g. oval or circular, cross section. Accordingly, the detection volume may have a frusto-conical or frusto-pyramidal shape.

[0089] When detecting moving insects in a field of vegetation, it has turned out that a frusto-conical or frusto-pyramidal detection volume having an elongated (e.g. elliptical or rectangular) base/cross section is particularly advantageous. The base/cross section may be a cross section orthogonal to an optical axis of the detector unit or of the illumination unit. In some embodiments, the elongated cross-section/base has a width (measured in a horizontal direction) that is larger than a height (measured in a vertical direction), e.g. such that the ratio between the width and the height is at least 3:2, such as at least 2:1, e.g. between 3:2 and 5:1, such as between 3:2 and 3:1, such as between 2:1 and 3:1. A detection volume having an elongated cross section with a horizontal longitudinal axis where the detection volume is elevated above the ground surface by a minimum vertical offset allows the detection volume to be arranged as a relatively flat volume, e.g. a flat box-shaped volume or a volume generally shaped as a flat pie slice, that may be horizontally arranged above a canopy of vegetation. Such a volume reduces reflections, stray light or other disturbing effects of the plants that might otherwise interfere with the detection process. Also, such a detection volume makes efficient use of the available illumination power to illuminate a volume where most insect activity occurs.

[0090] At least in some embodiments, the detection volume is a three-dimensional detection volume extending outside the insect sensor device. In particular, the detection volume is an enclosure-free void allowing unrestricted movement of living airborne insects into and out of the void. To this end, the detection volume may be defined solely by the overlap of the illumination volume and the field of view and depth of field of the detector module. When the detection volume is defined by an overlap between the illumination volume and the field of view and depth of field of the detector module, the illumination module may be configured to illuminate a conical or pyramidal or frusto-conical or frusto-pyramidal illumination volume, in particular with an elongated base/cross-section as described above with reference to the detection volume.

[0091] In some embodiments, the detector module comprises an imaging system, such as a camera. The imaging system includes an optical lens configured to image an image plane onto an image sensor, e.g. a quadrant Silicon detector or an image sensor having a lower or higher resolution. In some embodiments, the image plane is located between 1 m and 5 m, such as between 1 m and 4 m, such as between 1.5 m and 3 m in front of the optical lens. The imaging system is arranged such that the field of view of the imaging system overlaps, or even substantially coincides, with the illuminated volume at least at said image plane.

[0092] The imaging system may have a field of view and a depth of field large enough to record images of the entire detection volume, in particular sufficiently focused images to allow detection of the optically detectable attributes used for calculating the index of insect biodiversity. The imaging

system may be configured to detect disturbing events, e.g. larger animals or plants crossing the detection volume. The imaging system may also serve as a detector for detecting background radiation. In some embodiments, the sensor signals recorded by the image sensor may be used by the apparatus to detect insects and/or for detecting airborne trajectories of the insects, detecting wing beat frequencies and/or other attributes. The airborne trajectories are also examples of optically detected attributes that may serve as input to the computation of the index of insect biodiversity and/or another quantity associated with insect activity.

[0093] In some embodiments, the one or more detectors, e.g. the image sensor, comprise one or more photo diodes. Individual photodiodes that receive light from the entire detection volume or from a part of the detection volume allow for a fast time-resolved detection of changes in the intensity of backscattered light. Such signals may be used to determine wing beat frequencies of flying insects which, in turn, may be used to detect the presence of insects and, optionally, to distinguish between different types of insects based on properties of the wing beat patterns, e.g. the relative amplitudes of multiple frequencies in a frequency spectrum associated with a detected insect event.

[0094] In some embodiments, the detector module comprises an array of photodiodes, the photodiodes configured to receive light from respective parts of the detection volume, e.g. a linear or other 1D array or a 2D array, or another form of 2D image sensor that allows a spatially resolved detection of light impinging on different areas of an image plane. The detector module may be configured to direct light from different sub-volumes of the detection volume onto respective photo-diodes of the array or onto respective areas of a 2D image sensor, thus allowing a space-resolved detection of insects. In some embodiments, the array of photosensitive elements comprises no more than 128 photodiodes, such as no more than 64, such as no more than 25, such as no more than 16, such as no more than 9, such as no more than 4 photodiodes.

[0095] In some embodiments, the detector module is configured to selectively detect light at one or more predetermined wavelengths or one or more small wavelength bands. In some embodiments, the detector module is configured to selectively detect light at two or more wavelengths or small wavelength bands where the two or more wavelengths or wavelength bands are spaced apart from each other and do not overlap each other. To this end, the detector module may comprise one or more light-sensitive sensors—e.g. one or more photodiodes, photodiode arrays or other image sensors—configured to selectively detect light at two or more wavelengths or small wavelength bands where the two or more wavelengths or wavelength bands are spaced apart from each other and do not overlap each other. This may e.g. be achieved by a light-sensitive sensor where respective bandpass filters are selectively and alternatingly positioned in front of respective light-sensitive areas of the sensor. Alternatively, the detector module may include two or more separate light-sensitive sensors, each configured to detect light at a respective wavelength or wavelength band. Yet alternatively or additionally, the detector module may be configured to electronically separate sensor signals from light at the respective wavelength bands, e.g. by employing and filtering modulated light as described herein.

[0096] In particular, a detector module configured to selectively detect light at 808 nm and at 970 nm, respectively, has been found to be suitable for detecting and distinguishing different type of insects, e.g. based on a ratio of backscattered light at the respective wavelength. In some embodiments, the detector module comprises at least a first light-sensitive sensor configured to selectively detect light within a first wavelength band; and at least a second light-sensitive sensor to selectively detect light within a second wavelength band, non-overlapping with the first wavelength band.

[0097] Alternatively, the detector module comprises a light sensitive sensor configured to selectively detect light within a first wavelength band and within a second wavelength band, non-overlapping with the first wavelength band. In the latter embodiment, the detector module may include a modulation filter in order to separately register received light in the first and second wavelength bands and modulated at first and second modulation frequencies, respectively. Generally, the detector module may include a single detector or multiple detectors.

[0098] The detector module may be configured to obtain signals at a sampling rate of at least 1 MHz. A high sampling frequency (e.g. in the MHz range) facilitates the use of lock-in amplification which enables the ability to record two wavelengths on one diode and/or the suppression of background radiation.

[0099] Besides making it possible to distinguish signals at two wavelengths by a single detector, the modulation filter further ensures that only light with the right frequency and, optionally, the right phase can be detected by the detector module. This in turn ensures that the detector module is insensitive to light from other light sources such as the sun or other artificial sources. An efficient suppression of other light sources further ensures that detector data acquired at different locations and at different times are comparable and allow computation of a standardized biodiversity index.

[0100] In one preferred embodiment, the sensor signals are modulation filtered by means of a lock-in amplifier or by means of another suitable electronic modulation filter for extracting signals modulated at a target modulation frequency. When the illumination light at different wavelength bands is also modulated at different modulation frequencies, the detector module can separate sensor signals relating to different wavelength bands into separate channels based on the modulation filtering.

[0101] The processing unit may be configured to process the received sensor signals so as to detect one or more insect detection events and to extract one or more optically detectable attributes associated with the detected insect detection events. The processing may include one or more of the following: amplification, A/D conversion, filtering, calibration, feature detection, frequency analysis, calculation of attributes and/or the like.

[0102] In particular, the processing unit may process the sensor signal so as to detect one or more signal features indicative of the presence of one or more insects in the detection volume and extract from the sensor signal one or more optically detectable attributes associated with the detected insect detection events. The processing unit may further be configured to count the number of detected insect detection events, e.g. within a predetermined time period, a sliding window or the like, so as to determine an estimate of an amount of insects detected in the detection volume, e.g. as a number of insects detected in the detection volume, e.g. per unit time and/or per unit volume.

[0103] Accordingly, the processing unit may output processed detector data representing the respective insect detection events and the detected attributes.

[0104] In some embodiments, some or all of the processing steps are performed by a processing unit external to the insect sensor device, i.e. the processing steps may be implemented in a device external to the insect sensor device or they may be distributed between a local processing unit of the insect sensor device and a remote processing unit, separate from the insect sensor device.

[0105] For example, in such embodiments, the local processing unit of the insect sensor may output sensor data representing the detected sensor signals, optionally suitably pre-processed, and the external processing unit may further process the sensor signals so as to extract the optically detectable attributes. The external processing unit may be separate from a data processing system that performs the computation of the index of insect biodiversity and/or of another quantity associated with insect activity, or it may be integrated therein.

[0106] In some embodiments, the processing unit is configured to extract one or more optically detectable attributes associated with the detected insect detection events. The optically detected attributes may include one or more optically detectable attributes that can be determined from the sensor signals acquired by the optical insect sensor device. Examples of optically detectable attributes include: one or more wing beat frequencies, a body-to-wing ratio, a melanisation ratio (colour), a detected trajectory of movement of an insect inside the detection volume, a detected speed of movement of an insect inside the detection volume, an insect glossiness, or the like. In some embodiments, the optically detected attributes include a representation of light intensities associated with the insect detection event. The representation of light intensities may include a time-resolved and/or frequency-resolved representation, one or more features of a time-resolved and/or frequency-resolved representation, a processed version of a recorded time-resolved and/or frequency-resolved representation, and/or the like. For example, the representation may include time-resolved intensities at one or respective wavelength bands. Suitable features of a representation may include one or more locations of maxima and/or minima of the representation, one or more maximum or minimum values of the light intensity, locations, sizes and/or widths of one or more detected peaks in the representation and/or other detectable features. Examples of a processed version of a recorded time-resolved and/or frequency-resolved representation include a compressed version, an encoded version, an auto-encoded version and/or a dimensionally reduced version of the recorded time-resolved and/or frequency-resolved representation. In some embodiments, the computation of the index of insect biodiversity and/or of another quantity associated with insect activity is based on a combination of two or more optically detectable attributes. The detector data from each insect sensor may be indicative of an amount, e.g. a number, of detected insects detected in the detection volume during a sampling period. The detector data may include one or more optically detected attributes associated with each detected insect detection event and/or another suitable representation of the detected attributes, e.g. a distribution of attributes detected during a sampling period. It will be appreciated that, in some embodiments, the detector data may include unprocessed or only partially processed data, e.g. time-resolved detected light intensities or spectra from which one or more optically detectable attributes may be extracted or which itself may serve as optically detected attributes.

[0107] The detection and/or identification of insects based on wing beat frequencies, melanisation ratios and insect glossiness is described in more detail in WO 2017/182440 and in Gebru et. Al: "Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight", J. Biophotonics. 2018. While the above documents describe these indicators in the context of a LIDAR system using the Scheimflug principle and in the context of classification of insects, the present inventors have realized that attributes extracted by these techniques may also be applied to insect sensor devices based on other light sources that illuminate an extended volume rather than a narrow laser beam and to the computation of a biodiversity index instead of a taxonomic classification of insects. For example, WO 2017/182440 discloses a laser-based LIDAR system for detecting aerial fauna. Such a LIDAR system for aerial fauna utilizes a collimated laser beam that is transmitted relatively far into the atmosphere, and a receiver/detector measures the backscattered laser light from insects. While such an instrument is able to collect a large number of recordings, LIDAR systems are generally alignment sensitive and they require high-power lasers in order to provide a sufficiently long-range laser beam of sufficient intensity. Accordingly, such a system requires careful installation and operation of a high-power laser typically requires supervision and is normally not suitable for operation in e.g. urban areas.

[0108] Here and in the following, the term processing unit is intended to comprise any circuit and/or device suitably adapted to perform the functions described herein. In particular, the term processing unit comprises a general- or special-purpose programmable microprocessing unit, such as a central processing unit (CPU) of a computer or of another data processing system, a digital signal processing unit (DSP), an application specific integrated circuits (ASIC), a programmable logic arrays (PLA), a field programmable gate array (FPGA), a special purpose electronic circuit, etc., or a combination thereof.

[0109] In some embodiments the insect sensor device comprises or is communicatively coupled to one or more environmental sensor devices for sensing environmental data, such as weather data. Examples of environmental data include ambient temperature, humidity, amount of precipitation, wind speed, etc. The one or more environmental sensor devices may be included in the same housing as the optical sensor or it may be provided as a separate unit, e.g. a weather station, that may be communicatively coupled to an insect sensor device and/or to a remote data processing device. In some embodiments, a system of insect sensor devices may include one or more environmental sensor devices.

[0110] According to a fourth aspect, disclosed herein are embodiments of a system of insect sensors, the system comprising a plurality of optical insect sensor devices configured to be individually positioned within a geographic area, each insect sensor device configured to:

[0111] monitor insect activity within a three-dimensional detection volume extending outside the insect sensor device by detecting light from the detection volume, wherein the detection volume is an enclosure-

free void allowing unrestricted movement of living airborne insects into and out of the void; and to

[0112] output detector data indicative of one or more optically detected attributes associated with respective detected insect detection events, each insect detection event being indicative of one or more insects being present in the detection volume.

[0113] The system thus allows consistent and uniform monitoring of insect activity throughout even a large geographic area. Moreover, the monitoring process is performed with little or no disturbance to the natural insect behavior, thus avoiding undesired biases. The system is relatively cost effective and allows automatic operation and data collection.

[0114] The system may further comprise a data processing system communicatively coupled to the plurality of optical insect sensor devices and configured to:

[0115] receive detector data from respective ones of the plurality of optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective detected insect detection events, and to

[0116] compute, from the received detector data, one or more quantities indicative of insect activity within the geographic area, e.g. an index of biodiversity, the number of detected insects of respective insect species, and/or the like.

[0117] Each of the insect sensors may be an optical insect sensor according to the third aspect described above. Each insect sensor device may be mounted at a stationary detection site or non-stationary, e.g. mounted on a vehicle, e.g. as described above.

[0118] In some embodiments, the detector data from respective insect sensor devices is calibrated according to a detector reference that is uniform across the plurality of insect sensors.

[0119] Accordingly detector data from different insect sensor devices are comparable. In particular, detector data from different insect sensor devices may be used for the calculation of a single index of insect biodiversity and/or for the calculation of other quantities related to insect activity. Moreover, indexes of biodiversity (or other quantities) calculated from detector data from different insect sensor devices are directly comparable.

[0120] In some embodiments, the calibration according to a uniform detector reference includes one or more of the following:

[0121] 1) A uniform calibration of the time axis for the detected time-resolved light intensities, thereby extracting consistent frequency characteristics of the wing beat behaviour. This may be achieved by uniform calibration of the respective processor clocks of the insect sensor devices.

[0122] 2) A standardisation of the noise characteristics of the insect sensor devices in order for the insect sensor devices to have similar level of detection and in turn yield consistent body to wing ratios. The noise characteristics of the insect sensor devices are primarily determined by the characteristics of the light-sensitive sensor, e.g. by the characteristics of the quadrant detector and the performance of the transimpedance amplifier circuits associated with each of the individual detectors of the quadrant detector. These may be standardised by employing suitable manufacturing tolerances.

[0123] 3) Calibration/Standardisation of the detectivity of each insect sensor device for each wavelength, thereby providing consistent melanisation ratio detection of different sensors.

[0124] The detectivity of the insect sensor device for a given wavelength is a combination of many factors including:

[0125] 1) The level and exact spatial profile of the light emitted from the illumination module, in particular the spatial overlap of the two wavelengths throughout the entire detection volume.

[0126] 2) The exact alignment of the optical detector module including lens, quadrant detector and the relative alignment of the two.

[0127] As the detectivity is the sum of several interdependent variables, it is not always possible to ensure an exact match of both factors between insect sensor devices. Consequently, a standardisation process is advantageously used to measure offsets between insect sensor devices and/or between different sensor areas of the same insect sensor device so as to allow for a suitable calibration. In some embodiments, this may be achieved by a calibration process where spheres or other objects of respective predefined colours are dropped within the detection volume of an insect sensor device. This may be done across the entire detection volume. By monitoring the ratio of signals in the different wavelength channels of the detector module a characteristic ratio for each insect sensor device can be detected and subsequently compensated for, e.g. by determining one or more offsets and/or one or more multiplicative factors and/or one or more other calibration functions for adjusting one or more of the respective intensity levels in the one or more wavelength bands. In some embodiments, respective calibration functions may be determined for each light-sensitive area, e.g. for each photodiode of an array of light sensitive areas. Accordingly, in some embodiments, each insect sensor device is configured to output calibrated detector data, calibrated at least based on a set of wavelength-specific detectivity data indicative of a device-specific and wavelength-specific detectivity of the insect sensor device in respect of one or more predetermined calibration objects within the detection volume at respective wavelengths.

[0128] In some embodiments, each insect sensor device comprises a communications interface for transferring data on detected insect detection events from the insect sensor device to the data processing system so as to allow the data processing system to collect standardised data from a plurality of optical insect sensor devices.

[0129] The communications interface may be a wired or a wireless interface configured for direct or indirect communication of detector data to the data processing system. For example, indirect communication may be via a gateway device, via one or more other insect sensor devices or another node for relaying the detector data. The communication may be via a suitable communications network, such as via a cellular telecommunications network, e.g. using GSM/GPRS, UMTS, EDGE, 4G, 5G or any other suitable cellular telecommunications standard. In some embodiments, the communications interface may be configured for communication via satellite. Alternatively or additionally, the insect sensor device may include a local data storage device for logging the detector data and for allowing the stored data to be retrievable via a data port or a removable data storage device.

[0130] The data processing system may be implemented as one or more suitably programmed computers, such as a stand-alone computer, as a plurality of communicatively coupled computers, e.g. as client server-system, as a virtual computer or the like. The data processing system may directly or indirectly be communicatively coupled to the one or more insect sensor devices and receive the collected detector data from the one or more insect sensor devices. To this end, the data processing system may comprise a suitable wired or wireless communications interface, e.g. as described in connection with the communications interface of the insect sensor devices.

[0131] The data processing system is configured, e.g. by a suitable computer program, to process the received detector data from the plurality of insect sensor devices, e.g. to compute one or more indexes of insect biodiversity from the received detector data and/or to compute another quantities associated to insect activity.

[0132] For example, the data processing system may compute a respective local index of insect biodiversity for each of the one or more individual insect sensor devices. Each local index of insect biodiversity may thus be indicative of the local insect biodiversity in the environment around a corresponding individual insect sensor device. Additionally or alternatively, the data processing system may compute an overall index of biodiversity from detector data received from a plurality of insect sensor devices located within a larger geographic area. The overall index of insect biodiversity may thus be indicative of insect biodiversity in a larger geographic area.

[0133] It will be appreciated that the processing unit and/or the data processing system may be implemented as a client-server or a similar distributed system, where the data acquisition and, optionally, some signal processing, is performed locally in the insect sensor device, while other parts of the data processing tasks may be performed by a remote host system.

[0134] In some embodiments, each insect sensor device may have an associated environmental sensor device while, in other embodiments, the system may include more or fewer environmental sensor devices compared to the number of insect sensor devices. Accordingly, a process of computing an index of biodiversity and/or another quantity associated with insect activity may, in addition to the optically detected attributes, further base the computation on environmental data associated with the optically detected attributes. For example, the process may base the computation of the index of biodiversity and/or of another quantity associated with insect activity on one or more of the following parameters indicative of environmental conditions during the period where the optically detected attributes have been recorded: ambient temperature, time of day, precipitation, humidity, wind speed, etc. or a combination thereof. To this end, a mathematical model may not only receive the optically detected attributes as input but also one or more environmental data or other additional data, such as the time of day, time of year, etc. Yet further, in some embodiments the process may compute the index of biodiversity and/or other quantity from the optically detected parameters and compensate the computed biodiversity index and/or other quantity for the sensed environmental data, e.g. by applying a suitable correction factor or other correction function. Suitable correction functions or factors may be based on training data, e.g. in the form of a trained machine learning model, a look-up table and/or the like.

[0135] The present disclosure relates to different aspects including the apparatus, sensor device and system described above and in the following, corresponding apparatus, systems, methods, and/or products, each yielding one or more of the benefits and advantages described in connection with one or more of the other aspects, and each having one or more embodiments corresponding to the embodiments described in connection with one or more of the other aspects and/or disclosed in the appended claims.

[0136] According to another aspect, disclosed herein are embodiments of a computer-implemented method of determining an insect biodiversity index, the method comprising:

[0137] receiving detector data from one or more optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective insect detection events, detected by the one or more insect sensor devices, wherein the detected insect detection events are taxonomically unclassified;

[0138] computing an index of insect biodiversity directly from the optically detected attributes associated with the taxonomically unclassified detection events.

[0139] According to another aspect, disclosed herein are embodiments of a data processing system configured to perform steps of the method described herein. In particular, the data processing system may have stored thereon program code adapted to cause, when executed by the data processing system, the data processing system to perform the steps of the method described herein. The data processing system may be embodied as a single computer or as a distributed system including multiple computers, e.g. a client-server system, a cloud based system, etc. The data processing system may include a data storage device for storing the computer program and detector data. The data processing system may include a communications interface for receiving detector data.

[0140] According to another aspect, a computer program comprises program code adapted to cause, when executed by a data processing system, the data processing system to perform the steps of the method described herein. The computer program may be embodied as a computer-readable medium, such as a CD-ROM, DVD, optical disc, memory card, flash memory, magnetic storage device, floppy disk, hard disk, etc. having stored thereon the computer program. According to one aspect, a computer-readable medium has stored thereon a computer program as described herein.

[0141] Additional aspects, embodiments, features and advantages will be made apparent from the following detailed description of embodiments and with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0142] Preferred embodiments will be described in more detail in connection with the appended drawings, where

[0143] FIG. **1** shows a schematic view of an apparatus for determining an insect biodiversity index. The apparatus comprises a system of insect sensor devices.

[0144] FIG. **2** schematically illustrates an embodiment of a data processing system.

[0145] FIGS. **3** and **4** schematically illustrate embodiments of an insect sensor device.

[0146] FIG. **5** schematically illustrates an example of sensor signals from a detector module of an embodiment of an insect sensor device as described herein.

[0147] FIGS. **6** and **7** illustrate examples of detection volumes.

[0148] FIG. **8** shows a flow diagram of an example of a method for determining an index of insect biodiversity.

[0149] FIGS. **9**A and **9**B shows insect counts collected by conventional trapping methods from two fields with different degrees of biodiversity.

[0150] FIGS. **10**A and **10**B show corresponding histograms of wing beat frequencies collected for the two fields by an insect sensor device as described herein.

[0151] FIG. **11** shows an example of a calculated biodiversity prediction based on sensor data for the low biodiversity field (left) and the high biodiversity field (right).

DETAILED DESCRIPTION

[0152] FIG. **1** shows a schematic view of an apparatus for determining an insect biodiversity index. The apparatus, generally designated by reference numeral **100**, comprises a system of insect sensor devices. In particular the apparatus comprises a data processing system **200** and a plurality of insect sensor devices **120**. The insect sensor devices are deployed throughout a geographic area **300**. Each insect sensor device **120** may be mounted on a suitable stationary or movable support, e.g. on a frame, on a stand, on a vehicle, etc.

[0153] As will be described in greater detail below, each insect sensor device may comprise an illumination module including a light source, such as one or more halogen lamps, one or more LEDs or the like, configured to illuminate an illuminated volume in a proximity of the insect sensor device. The insect sensor device may further comprise a detector module including one or more detectors and one or more optical elements configured to capture backscattered light from at least a portion of the illuminated volume and to guide the captured light onto the one or more detectors. The illuminated volume from which light is captured by the detector module for detecting insects is referred to as detection volume **150**. Generally, the detection volume may be defined as the volume from which the detector module obtains light signals useful for detecting insects. The detection volume is typically defined by an overlap of the volume illuminated by the illumination module and by the field of view and depth of field of the detector module. In particular, the detection volume is not limited by any physical enclosure but is an open, unenclosed void or space which airborne, living insects may enter or exit in an unrestricted manner.

[0154] The insect sensor device comprises a processing unit configured to perform the detection of insects and to forward information about the detected insect population and associated optically detectable attributes to the data processing system.

[0155] Preferably, the insect sensor devices **120** are of the same type and calibrated and standardized according to a common detector reference, thus allowing detector data and attributes determined by them to be compared with each other and/or to be used as input to a computation of an overall index of insect diversity and/or another quantity indicative of insect activity associated with the geographic area **300**.

[0156] The data processing system **200** may be a standalone computer or a system of multiple computers, e.g. a client-server system, a cloud-based system or the like. An example of a data processing system will be described in more detail below with reference to FIG. **2**.

[0157] Each insect sensor device **120** is an optical insect sensor device using reflected/backscattered light from insects in a detection volume **150** to detect insects and to measure optically detectable attributes of the detected insects, e.g. one or more of the following: one or more wing beat frequencies, a body-to-wing ratio, a melanisation ratio (colour), a detected trajectory of movement of an insect inside the detection volume, a detected speed of movement of an insect inside the detection volume, an insect glossiness, or the like.

[0158] Generally, the insect sensor device detects insect detection events. An insect detection event refers to the detection of one or more insects being present in the detection volume. Detection of an insect detection event may be based on one or more trigger criteria, e.g. based on a signal level of the detected sensor signal and/or on another property of the sensor signals sensed by the detector module of the insect sensor device in response to the received light from the detection volume.

[0159] The detection volume **150** associated with each insect sensor device is a detection volume external to the corresponding insect sensor device located in the vicinity of the insect sensor device. An example of an insect sensor device will be described in more detail below with reference to FIG. **3**. Examples of detection volumes will be described in more detail below with reference to FIGS. **6** and **7**.

[0160] In the example of FIG. **1**, the apparatus comprises three like insect sensor devices. It will be appreciated that other embodiments may include fewer or more insect sensor devices. For example, some embodiments may only include a single insect sensor device, while other embodiments may include 5, 10, 20, 100 or even more insect sensor devices. It will be appreciated, that the number of insect sensor devices may be chosen depending on factors such as the size and variability of the geographic area, the desired accuracy of the resulting biodiversity index or other computed quantity, the spatial resolution of respective local indexes of biodiversity or other local quantities, etc.

[0161] Each insect sensor device **120** is communicatively coupled to the data processing system **200** and communicates the collected detector data, including measured attributes to the data processing system **200**. In the example of FIG. **1**, each insect sensor device communicates the collected detector data via a cellular telecommunications network to the data processing system **200**, e.g. via a GSM/GPRS network, USTM network, EDGE network, 4G network, 5G network or another suitable telecommunications network. It will be appreciated that the communication may be a direct communication or via one or more intermediate nodes. Similarly, the communication may use alternative or additional communications technologies, e.g. other types of wireless communication and/or wired communication. Yet further, the collected detector data may be stored locally by the insect sensor device for subsequent manual retrieval from each insect sensor device, e.g. on a portable data storage device and subsequent input to the data processing system **200**.

[0162] The data processing system **200** is configured to execute a computer program for analysing the detector data

from one or more insect sensor devices and for computing one or more desired quantities indicative of insect activity. In particular, the data processing system **200** may be configured to compute an index of insect biodiversity directly from the detector data as described herein, i.e. without intermediate taxonomic classification of the detected insects. The data processing device may output the computed index of insect biodiversity or other computed quantity in a suitable form, e.g. on a display, on a storage device, via a data communications interface, and/or the like.

[0163] In the example of FIG. **1**, one of the insect sensor devices comprises an environmental sensor device **180** for sensing one or more environmental parameters, such as temperature, wind speed, humidity and/or other weather data. The sensed environmental data is also communicated to the data processing system, e.g. directly by the environmental sensor device or by the insect sensor device. It will be appreciated that, in some embodiments, each insect sensor includes or is operationally coupled to, an environmental sensor device. In other embodiments only a subset of insect sensor devices include or are operationally coupled to an environmental sensor device. Alternatively or additionally, the apparatus may include one or more environmental sensor devices deployed in the geographic area **300** and communicatively coupled to the data processing system **200**, separately from the insect sensor devices. The data processing system may thus also base the computation of the index of biodiversity or other quantity on the environmental data sensed by the environmental sensor device(s), e.g. by computing a modified index of insect biodiversity or otherwise relating the computed index of insect biodiversity to the sensed environmental conditions.

[0164] It will be appreciated that, while the system of the plurality of insect sensor devices and the data processing system of FIG. **1** is particularly suitable for determining an index of biodiversity, embodiments of the system may also be used to determine other quantities indicative of insect activity within a geographic area, e.g. for monitoring the insect activity within an agricultural production area, such as one or more fields for growing crops.

[0165] FIG. **2** shows a schematic view of an example of a data processing system **200**, e.g. the data processing system **200** of the apparatus of FIG. **1**.

[0166] The data processing system **200** comprises a central processing unit **240** or other suitable processing unit. The data processing system further comprises a data storage device **230** for storing program code, received detector data and, optionally, a mathematical model for computing the index of insect biodiversity. Examples of suitable data storage devices include a hard disk, an EPROM, etc. The data processing system further comprises a data communications interface **270**, e.g. a network adaptor, a GSM module or another suitable circuit for communicating via a cellular communications network or via another wireless communications technology. To this end, the data processing system further comprise an antenna **271**. It will be appreciated that the data processing system may include a wired data communications interface instead of or in addition to a wireless communication interface. The data processing system further comprises an output interface **220** e.g. a display, a data output port, or the like.

[0167] FIG. **3** schematically illustrates an embodiment of an insect sensor device, e.g. one of the insect sensor devices of the system of FIG. **1**. The insect sensor device, generally designated by reference numeral **120**, comprises a processing unit **140**, a detector module **130** and an illumination module **131**, all accommodated within a housing **110**. In this example, the illumination module and the detector module are vertically aligned with each other and the illumination module is arranged below the detector module. However, other arrangements are possible as well.

[0168] Generally, in order to maximize the amount of backscattered light from insects inside the detection volume **150**, it may be preferable to position the illumination module adjacent or otherwise close to the detector module, such that the illumination direction and the viewing direction only define a relatively small angle between them, e.g. less than 30°, such as less than 20°. In some embodiments, the illumination module is configured to emit a beam of light along an illumination direction, and the detector module defines a viewing direction, e.g. as an optical axis of the detector module, wherein the illumination direction and the viewing direction define an angle between each other, the angle being between 1° and 30°, such as between 5° and 20°.

[0169] The illumination module comprises an array of light-emitting diodes (LEDs) **161** and a corresponding array of lenses **162** for directing the light from the respective LEDs as a diverging beam **163** along an illumination direction **164**. The array of light emitting diodes may comprise a first set of diodes configured to selectively emit light at a first wavelength band, e.g. at 808 nm+/−25 nm. The array of light emitting diodes may further comprise a second set of diodes configured to selectively emit light at a second wavelength band, different from the first wavelength band, in particular spaced-apart from the first wavelength band, e.g. at 970 nm+/−25 nm. In other embodiments, the array of light emitting diodes may include alternative or additional types of LEDs or only a single type of LEDs. For example, in some embodiments, the LEDs may be configured to emit broad-band visible, near-infrared and/or infrared light.

[0170] The detector module **130** comprises an optical system **132** in the form of a Fresnel lens. Alternative another lens system may be used, e.g. an NIR coated aspheric lens, e.g. having 60 mm focal length and an ø76.2 mm aperture. The detector module **130** includes an optical sensor **133**, e.g. one or more photodiodes, such as an array of photodiodes, a CCD or CMOS sensor and the optical system directs light from the detection volume onto the optical sensor. In some embodiments, the optical system images an object plane **152** inside the illuminated volume onto the optical sensor. The field of view of the optical system and the depth of field of the optical system are configured such that the optical system directs light from a portion of the volume illuminated by the illumination module onto the optical sensor. The portion of the illuminated volume from which the optical system receives light such that it can be detected by the optical sensor and used for detection of insects defines a detection volume **150**. The optical system **132** defines an optical axis **134** that intersects with the illumination direction **164**, preferably at a small angle, such as 10°.

[0171] For example, when an optical system is based on a camera lens having f=24 mm, f/2.8 and an optical sensor includes a ¾" image sensor, the detector module may be configured to focus on an object plane at 2 m distance from the lens, corresponding to a field of view of approximately 1.7 m×1.7 m and a depth of field of approximately 1.3 m, thus resulting in a detection volume of approx. 3.7 m^3.

[0172] The detector module **130** is communicatively coupled to the processing unit **140** and forwards a sensor signal indicative of the captured radiation by the optical sensor **133** to the processing unit. The processing unit **140** may include a suitably programmed computer or another suitable processing device or system. The processing unit receives the sensor signal, e.g. an image or stream of images and/or one or more sensed light intensities from respective one or more photodiodes and, optionally, further sensor signals from the detector module. The processing unit **140** processes the received sensor signals so as to detect and identify insects in the detection volume and output detector data indicative of detected insect detection events and associated optically detectable attributes.

[0173] FIG. **4** schematically illustrates a more detailed view of an example of the insect sensor device. The insect sensor device **120** of FIG. **4** is similar to the insect sensor device of FIG. **3** and comprises a processing unit **140**, a detector module **130** and an illumination module **131**, all accommodated within a housing **110** and all as described in connection with FIG. **3**. In this example, the illumination module **131** includes an array of light emitting diodes (LEDs). The LEDs may be arranged in a 2D pattern, such as on a regular 2D grid. The LEDs may be distributed over an area of at least 10 cm^2, such as at least 30 cm^2, such as at least 60 cm^2, such as at least 80 cm^2. In some embodiments, the LEDs may be distributed over an area between 10 cm^2 and 400 cm^2, such as between 30 cm^2 and 300 cm^2, such as between 40 cm^2 and 200 cm^2, such as between 60 cm^2 and 120 cm^2, e.g. about 90 cm^2. Accordingly, an illumination beam having a large cross-sectional area may be emitted so as to illuminate a large volume simultaneously. The light emitted from each diode may be partially collimated by an asymmetrical lens to form a diverging beam, e.g. expanded with 40° and 8° diverging angles in the vertical and horizontal axis, respectively (measured as full divergence angles). The array of LEDs may all emit the same wavelength band or be arranged in such a way as to mix multiple wavelengths. In one example, the illumination module emits light at two different narrow wavelength bands, i.e. a first band at a first wavelength and a second band at a second wavelength, such as at 808 nm and 970 nm, respectively. Other embodiments may include a single type of LEDs or more than two different types of LEDs. The light from the illumination module is modulated at one or at multiple respective frequencies, e.g. the light at each wavelength may be encoded with a unique frequency. In one example, the light at the first wavelength is modulated at a first modulation frequency and the light at a second wavelength is modulated at a second modulation frequency, different form the first modulation frequency. The first and second modulation frequencies may each be selected between 10 kHz and 500 kHz, such as between 50 kHz and 200 kHz. In one example, the first modulation frequency is about 80 kHz and the second modulation frequency is about 120 kHz. To this end, the processing unit includes a synchronization circuit **141** having a clock for controlling the illumination module.

[0174] The detector module **131** includes an image sensor **133** including a 2×2 array of light-sensitive elements, such as photodiodes. In one particular embodiment, the image sensor is a quadrant detector with four individual Si photodiodes arranged in a square. It will be appreciated that other embodiments may include a larger array of light-sensitive elements or a smaller array or light sensitive elements, such as a 2×1 array, or even a single light sensitive element. The optical system **132** is arranged relative to the photodiode sensor array in such a way as to image an image plane within the detection volume onto the photodiode array. The four light-sensitive areas thus collect light from four substantially separate sub-volumes of the detection volume.

[0175] The detected signals from the photodiode array **133** are fed into the processing unit **140**. The processing unit includes an amplifier bank **142** with a number of amplifiers matching the size of the photodiode array. In this example, the amplifier bank includes four transimpedance amplifiers. The amplified signals are fed into a corresponding A/D converter bank **143** which includes a number of A/D converters corresponding to the size of the photodiode array, such as four A/D converters. The A/D converter bank **143** generates respective digital time-resolved signals for the individual photodiodes. The processing unit further comprises a de-multiplexer circuit **144**, e.g. an FPGA implementing a number of digital lock-in amplifiers corresponding to the size of the photodiode array and to the number of wavelengths. In one example, the de-multiplexer circuit implements eight lock-in amplifiers corresponding to the four quadrants of the quadrant detector and two individually modulated wavelengths. The de-multiplexer circuit **144** demultiplexes the signals from each of the photodiodes into separate signals, optionally into separate signals for the respective wavelengths, i.e. for each photodiode, the demultiplexer circuit generates one signal for each individually modulated wavelength. To this end, the de-multiplexing circuit receives a clock signal from the synchronisation circuit **141**. The lock-in amplifiers further serve as an efficient filter for light not modulated with frequencies around the two lock-in frequencies.

[0176] The resulting de-multiplexed signals thus include one or more, e.g. two, wavelength-specific channels for each photodiode, e.g. 2×4 channels. It will be appreciated that, in embodiments with a different number of wavelengths or a different array size, the number of de-multiplexed signals will generally be different. The de-multiplexed signals are forwarded to a data processing circuit **145** which processes the individual signals to detect insects being present in the detection volume, i.e. to detect insect detection events, and to determine one or more attributes of each detected insect. To this end, the data processing circuit **145** may initially perform a calibration of the signal, e.g. based on stored calibration data, such as stored offsets and/or multiplicative factors. The data processing circuit outputs detector data indicative of the insect detection events and the associated determined attributes. The data processing circuit may further log detector data associated with multiple insect detection events. The data processing circuit may intermittently, e.g. periodically, upon request, or when the internal log buffer is about to be full, communicate the recorded detector data via the communications interface **170** to a remote data processing system as described herein.

[0177] FIG. **5** schematically illustrates an example of de-multiplexed sensor signals from a detector module of an embodiment of an insect sensor device as described herein, e.g. an insect sensor device as described in connection with FIG. **3** or **4**. In this example, the sensor signals from the detector module includes respective time series of detected light intensities at two narrow wavelength bands, e.g. as recorded by respective photodiodes provided with respective bandpass filters or by one of the photodiodes of the array

of FIG. **4**. In some embodiments the signal may be integrated or otherwise combined from multiple photodiodes, from an image sensor and/or the like.

[0178] In this example, time series **701** corresponds to detected light at 808 nm while time series **702** corresponds to detected light at 970 nm. However, other embodiments may use other wavelengths and/or more than two wavelengths or wavelength bands.

[0179] The processing unit of an insect sensor device may process the times series to detect the presence of an insect in the detection volume and to determine one or more attributes of the detected insect. Alternatively, some or all of the signal and data processing may be performed by a data processing system external to the insect sensor device.

[0180] In the present example, the process implemented by the processing unit and/or an external data processing system may detect the presence of detected radiation above a predetermined threshold and/or determine a fundamental harmonic of the detected frequency response so as to detect the presence of an insect, i.e. to identify an insect detection event.

[0181] For example, in one embodiment, the processing unit of the insect sensor device records data for a given interval (typically ten minutes), extracts events and metadata and then starts a new recording. The recorded data may include respective time series of the de-multiplexed channels of sensor signals.

[0182] To extract the events from the recorded raw data, the process estimates a rolling temporal mean and standard deviation. To this end, in each window, the data is reduced by a factor **10** before the mean and standard deviation is calculated.

[0183] An event threshold is then defined by multiplying the estimated standard deviation with a signal to noise factor (SNR), resulting in a threshold map representing the data of the respective channels.

[0184] Finally, the estimated rolling mean is removed from the signal and the events are extracted by applying the threshold map. The data associated with the extracted events are stored on the insect sensor device and uploaded, e.g. via cellular connection, to a cloud database or other suitable data repository as soon as a connection is available. In cases where no cellular or other data connection is available, it is possible to store the extracted events locally on the insect senor device.

[0185] A process implemented by a cloud service or another type of data processing system external to the insect sensor device may perform data processing of the recorded data associated with the detected insect detection events. It will be appreciated, however, that some or even all of the subsequent processing may also be performed locally on the insect sensor device.

[0186] In any event, the process may compute one or more attributes of the insects associated with the detected insect events. Examples of such attributes include a fundamental wing beat frequency (WBF), a body-wing ratio (BWR) and a melanisation ratio (MEL).

[0187] For example, the process may compute the fundamental wing beat frequency (WBF) from the determined fundamental harmonic of the frequency response of a detected detection event. The process may compute the body-wing ratio as a mean ratio between a wing and body signal. The body signal may be determined as a baseline signal **711** of a detection event which represents the scattering from the insect with closed wings while the wing signal may be determined as the signal levels **712** at the peaks in scattering.

[0188] The melanisation ratio may be determined as a mean ratio between the signal strengths of the two recorded channels during a detection event.

[0189] Based on respective sets of one or more of the above attributes, associated with a plurality of insect detection events, optionally in combination with other parameters, a data processing system may compute an index of insect of biodiversity.

[0190] Generally, embodiments of the insect sensor device described herein provide a detection volume that is large enough for the detector module to observe a number of insects representative for the population density in the area. The detection volume is also small enough to be sufficiently uniformly illuminated so as to provide high signal strength at the image sensor.

[0191] Moreover, embodiments of the apparatus described herein provide fast observation times, e.g. so as to reliably detect insects even in situations of high insect activity. Moreover embodiments of the apparatus described herein provide long enough observation times to be able to reliably determine attributes of the flying insects.

[0192] FIGS. **6** and **7** illustrate examples of detection volumes. FIG. **6** schematically shows an example of a frusto-conical detection volume resulting from an illumination module emitting a diverging light beam with a generally circular cross section. FIG. **7** schematically illustrates an example of a frusto-pyramidal detection volume.

[0193] In order to compute an accurate index of insect biodiversity it is preferable that the recorded insect activity is representative for the area under consideration. In order to achieve this, sufficiently high counting statistics are preferred.

[0194] As described herein, some embodiments of the insect sensor device described herein record one or more time series of light scattering off one or more insects in flight at one or more wavelengths of the light. From the recorded time series, the wing beat frequency and/or ratio of scattering from body and wings, respectively, can be computed. However, in order to obtain a reliable and accurate detection result, the recorded time series should be long enough for multiple wingbeats to occur. The wingbeat frequency of insects in flight spans from around 10 Hz to around a 1000 Hz. In order to get more than 10 wings beats the time the insect is in the detection volume should, in the worst case, be preferably more than 100 ms or even 1 s. Similarly, a detection based on recorded flight trajectories is facilitated by observation times long enough to record trajectories of sufficient lengths.

[0195] Embodiments of the insect sensor device described herein thus employ a detection volume shaped and sized to allow sufficiently long observation times, even when sensor is moving across an area of land.

[0196] The extent of the detection volume in a direction along an optical axis of the detector module should preferably be larger than 50 cm, such as larger than 1 m, such as larger than 2 m, such as larger than 5 m in order to ensure that insects are likely to remain inside the detection volume sufficiently long. For example, the length of the detection volume along the optical axis of the detector volume may be less than 100 m, such as less than 50 m, such as less than 20 m, such as less than 10 m.

[0197] Furthermore, as discussed above, it is preferred that the total detection volume is of the order of, or larger than, 1 m^3 such as larger than 1 m^3. In order to achieve such a detection volume with one or a system of small and cost-efficient insect sensor devices, it is preferred that the illumination module is carefully configured, and that the detection volume of the individual insect sensor is relatively large, such as larger than 10 l.

[0198] The illuminated detection volumes shown in FIGS. **6** and **7** both provide large detection volumes in the vicinity of the insect sensor device, i.e. allowing representative and local measurements.

[0199] The detection volumes shown in FIGS. **6** and **7** represent an overlap between an illuminated volume, illuminated by an illumination module of the insect sensor device, and by a detectable volume from which a detector of the insect sensor device receives light, i.e. the detectable volume may be defined by a field of view and depth of field of the detector. In one embodiment, the illumination module comprises one or more suitable light sources, e.g. one or more high-power LEDs, emitting light which is diverging from the illumination module so as to distribute light into a large volume. In one particular embodiment, the illumination module is configured to emit light with a full divergence angle in the horizontal plane that is larger than 5°, such as larger than 10° such as larger than 20°, while the vertical divergence is limited to angles smaller than 2° such as smaller than 5°. The resulting detection volume consequently will be optimized in space just above the vegetation canopy. Moreover, in this embodiment, the amount of light which disappears upwards or into the vegetation is limited. In another embodiment, the illumination module is configured to emit light with a full divergence angle in the vertical plane that is larger than 5°, such as larger than 10° such as larger than 20°, such as larger than 30°, while the horizontal divergence is limited to angles smaller than 15° such as smaller than 10°. This embodiment allows a compact design of the insect sensor device with the detector and illumination modules arranged one above the other.

[0200] It is further preferred that the illumination module is configured so as to direct the illumination light along a center optical axis of the radiated light (i.e. along a direction of illumination) that points upwards in such an angle as to completely eliminate light form hitting the crop, e.g. between 1° and 30°, such as between 2° and 30°, such as between 5° and 20°.

[0201] An example of a detection volume resulting from such a diverging, pie-shaped, forward-upwardly directed illumination beam is illustrated in FIG. **7**. In particular, FIG. **7** illustrates a 3D view of the detection volume **150** as well as a side view and a top view of the detection volume. In the example of FIG. **7**, the distance d_0 between the aperture of the detector module and the start of the detection volume is about 1 m. The distance d_1 between the aperture of the detector module and the far end of the detection volume is about 10 m. The divergence angle $\theta_{vertical}$ of the diverging light beam in the vertical direction (full angle) is about 4° while the divergence angle $\theta_{Horizontal}$ in the horizontal direction (full angle) is about 20°. However, it will be appreciated that other embodiments may have different size and/or shape. For example, the divergence angle $\lambda_{vertical}$ of the diverging light beam in the vertical direction (full angle) may be about 40° while the divergence angle $\lambda_{Horizontal}$ in the horizontal direction (full angle) is about 8°.

[0202] Generally, when the detection volume is positioned close to the insect sensor device efficient illumination of the detection volume and reliable detection of small insects is facilitated. For example, the boundary of the detection volume closest to an aperture of the detector module may be between 1 cm and 10 m away from the aperture of the detector module, such as between 10 cm and 5 m, such as between 10 cm and 2 m. The boundary of the detection volume furthest from an aperture of the detector module may be between 3 m and 100 m away from the aperture of the detector module, such as between 5 m and 20 m, such as between 8 m and 12 m.

[0203] FIG. **8** shows a flow diagram of an example of a method for determining an insect biodiversity index spraying insecticides.

[0204] In initial step S**1**, a set of insect sensor devices are provided. In particular, the insect sensor devices are calibrated and standardized according to a common reference as described herein.

[0205] For example, the calibration process may include the following steps:

[0206] A plurality of objects, e.g. balls, of respective colors are caused to move in the detection volume, e.g. by dropping the objects such that they traverse the detection volume of the insect sensor device. The sensor signals for each of the objects are recorded. This is repeated a plurality of times with objects moving at different positions within the detection volume. Preferably, for each position, sensor signals for a plurality of object movements, such as at least 100, such as more than 500 movements of objects having different color are recorded. The sensor signals may represent detected intensities (denoted $I_{80\ 1}$ and $I_{\lambda 2}$, respectively) at two wavelength bands around wavelengths λ_1 and λ_2, respectively, as described herein. For each object, the ratio $I_{\lambda 1}/(I_{80\ 1}+I_{\lambda 2})$ or another measure of the relative detected intensity at the two wavelengths is calculated. In one embodiment λ_1=808 nm and λ_2=970 nm. In case of a sensor array such as a quadrant detector this may be done for each light-sensitive area of the array. The distribution of detected ratios may be recorded (e.g. for each light-sensitive array) and the sensor signals may be adjusted by respective calibration functions so as to cause the calibrated distributions to conform with a reference distribution, e.g. such that the peak of the distribution is located at a common reference value. Knowing these calibration functions, melanisation ratios observed on insects by different insect sensor devices can be compensated to yield identical results compensated for variations in spatial overlap of the wavelengths and the exact overlap of the two wavelengths.

[0207] In step S**2**, the insect sensors are deployed in a geographic area for which an index of insect biodiversity is to be computed.

[0208] In step S**3**, detector data indicative of respective insect detection events are recorded and associated optically detectable attributes are computed. This step may be repeated for a predetermined period of time or until sufficient data has been collected.

[0209] In step S**4** a biodiversity index is calculated based on detector data from the set of insect sensor devices. Again the collection of data and computation of a biodiversity index may be repeated, e.g. in order to analyse a change of biodiversity over time.

Example of Biodiversity Calculation—Total Variance Method

[0210] Insect biodiversity is conventionally measured by a combination of species richness (number of different species recorded in a sample) and aggregate statistics such as the Simpson's biodiversity index, which takes into account the relative abundance of species.

[0211] FIGS. **9**A and **9**B show insects collected by conventional trapping methods from two fields with different degrees of biodiversity. The figures show histograms representing the number of counts per species for two fields with different degrees of biodiversity. FIG. **9**A shows the number of counts for a field with low biodiversity while FIG. **9**B shows corresponding counts for a field with high biodiversity.

[0212] The insects have been taxonomically classified by an expert in the field using a microscope. In the low biodiversity field, a total of 14 different insect species were identified, whereas in the high biodiversity field 25 different insect species were identified over the same period in the same traps. Including the relative numbers of each species, a Simpson's biodiversity index can be calculated of 0.635 for the low biodiversity field and 0.828 for the high biodiversity field.

[0213] FIGS. **10**A and **10**B show corresponding histograms of wing beat frequencies collected for the two fields by an insect sensor device as described herein. FIG. **10**A shows the number of counts for a field with low biodiversity while FIG. **10**B shows corresponding counts for a field with high biodiversity. From visual inspection it is clear that there are wingbeat frequencies present in the high biodiversity field that are not present in the low biodiversity field, particularly around the higher frequencies between 200 Hz and 300 Hz.

[0214] Both fields have a dominant peak at 100 Hz, but this peak is slightly less dominant in the high biodiversity histogram than the low biodiversity histogram.

[0215] Similar histograms are available with different features of field insects, such as body/wing ratio and melanisation, and each of these and their combination can be interpreted to provide a comparative picture of biodiversity.

[0216] Some embodiments of the apparatus described herein characterize, without the need for taxonomic identification of every detected insect, the insect diversity as a statistic measure computed from the distribution of attributes such as wing beat frequency.

[0217] FIG. **11** shows a calculated biodiversity prediction based on sensor data for the low biodiversity field (left) and the high biodiversity field (right).

[0218] Although the invention has been described with reference to certain specific embodiments, various modifications thereof will be apparent to those skilled in art without departing from the spirit and scope of the invention as outlined in claims appended hereto.

[0219] In summary, advantages of some or all of the disclosed embodiments include:

[0220] Real time monitoring

[0221] The ability to detect variations in biodiversity on a day to day, hour to hour and even minute to minute basis

[0222] Unbiased, no use of pheromones or other methods of manipulating insect activtiy are needed.

[0223] Cheap and accessible

[0224] Labour and "laboratory" free

[0225] Standardises data collection and format

[0226] Improves ability to compare biodiversity measurements

[0227] Resulting index may contain more information

[0228] Reduces risk of human and systematic error

[0229] It will be appreciated that insects vary a lot in size and behavior. Insect sizes can vary from less than one mm to a few cm and movement patterns of insects can vary from insects standing still, hovering, in air to jumping insects with ballistic trajectories. Embodiments of the apparatus and insect sensor device described herein have been found useful for various types of airborne insects, including flying insects having wings and jumping insects, such as jumping flea beetle, e.g. cabbage stem flea beetle (*Psylliodes chrysocephala*).

1-24. (canceled)

25. An apparatus for determining an index of insect biodiversity, comprising:

a plurality of optical insect sensor devices configured to be individually positioned within a geographic area, each of the plurality of optical insect sensor devices configured to:

monitor insect activity within a detection volume extending outside the insect sensor device by detecting light from the detection volume, and

output detector data indicative of one or more optically detected attributes associated with respective detected insect detection events, each insect detection event being indicative of one or more insects being present in the detection volume;

a data processing system communicatively coupled to the plurality of optical insect sensor devices and configured to:

receive the detector data from respective ones of the plurality of optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective detected insect detection events, and

compute, from the received detector data, an index of insect biodiversity indicative of insect biodiversity within the geographic area.

26. An apparatus according to claim **25**, wherein the detected insect detection events are taxonomically unclassified and wherein the data processing system is configured to compute the index of insect biodiversity directly from the optically detected attributes associated with the taxonomically unclassified detection events.

27. An apparatus for determining an index of insect biodiversity, comprising:

one or more optical insect sensor devices configured to be individually positioned within a geographic area, each of the one or more insect sensor devices configured to:

monitor insect activity within a detection volume to detect one or more insect detection events, each insect detection event being indicative of one or more insects being present in the detection volume, and

output detector data indicative of one or more optically detected attributes associated with respective detected insect detection events,

a data processing system communicatively coupled to the one or more of optical insect sensor devices and configured to:

receive the detector data from the one or more optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective detected insect detection events, wherein the detected insect detection events are taxonomically unclassified, and

compute an index of insect biodiversity directly from the optically detected attributes associated with the taxonomically unclassified detection events.

28. An apparatus according to claim **27**, wherein the data processing system is configured to compute the index of insect biodiversity as a measure of variability of the optically detected attributes associated with respective detected insect detection events.

29. An apparatus according to claim **28**, wherein the data processing system is configured to compute the measure of variability by performing a statistical analysis of the optically detected attributes, the statistical analysis being a multi-variate analysis associated with respective detected insect detection events.

30. An apparatus according to claim **28**, wherein the data processing system is configured to compute the measure of variability at least by performing a clustering of the detected insect detection events according to at least the optically detected attributes associated with the respective detected insect detection events.

31. An apparatus according to claim **27**, wherein the data processing system is configured to compute the index of insect biodiversity based on a mathematical model, the mathematical model being a trained machine learning model trained by supervised learning that directly maps the optically detected attributes associated with respective detected insect detection events to an index of insect biodiversity.

32. An apparatus according to claim **27**, wherein each insect sensor device of the plurality of insect sensor devices is calibrated based on a common detector reference.

33. An apparatus according to claim **27**, wherein each insect sensor device is configured to output calibrated detector data based on a set of wavelength-specific detectivity data indicative of a wavelength-specific detectivity of each insect sensor device in respect of one or more predetermined calibration objects within the detection volume at respective wavelengths.

34. An apparatus according to claim **27**, wherein the optically detected attributes include one or more attributes selected from the group of:

a detected trajectory of movement of an insect inside the detection volume;

a detected speed of movement of an insect inside the detection volume;

one or more detected wing beat frequencies;

a melanisation ratio;

an insect glossiness.

35. An apparatus according to claim **27**, further comprising one or more environmental sensor devices for sensing environmental data indicative of an environmental condition of the geographic area; and wherein the data processing system is configured to receive the environmental data and to compute the index of insect biodiversity from the received detector data and from the sensed environmental data.

36. An apparatus according to claim **27**, wherein each optical insect sensor device comprises:

an illumination module configured to illuminate the detection volume with illumination light comprising light at a first wavelength band modulated at a first modulation frequency;

a detector module comprising a detector configured to detect light from the detection volume; and

a processing unit configured to receive sensor signals from the detector module and configured to filter the received sensor signal to extract a first sensor signal modulated at the first modulation frequency and, based on the first sensor signal, to detect at least one insect in the detection volume and to determine at least one optically detectable attribute of the detected insect.

37. An insect sensor device for detecting insects in a detection volume, the insect sensor device comprising:

an illumination module configured to illuminate the detection volume with illumination light comprising light at a first wavelength band modulated at a first modulation frequency;

a detector module comprising a detector configured to detect light from the detection volume; and

a processing unit configured to receive sensor signals from the detector module and configured to filter the received sensor signal to extract a first sensor signal modulated at the first modulation frequency and, based on the first sensor signal, to detect at least one insect in the detection volume and to determine at least one optically detectable attribute of the detected insect.

38. An insect sensor device according to claim **37**, wherein the illumination light further comprises light at a second wavelength band, different from the first wavelength band, modulated at a second modulation frequency, different from the first modulation frequency, wherein the processing unit is further configured to filter the received sensor signal to extract a second sensor signal modulated at the second modulation frequency and, based on the first and second sensor signals, to detect at least one insect in the detection volume and to determine at least one optically detectable attribute of the detected insect.

39. An insect sensor device according to claim **37**, wherein the illumination module comprises an array of light-emitting devices.

40. An insect sensor device according to claim **39**, wherein the illumination module comprises a corresponding array of lenses for directing the light from the respective light-emitting devices as a diverging beam along an illumination direction.

41. An insect sensor device according to claim **37**, wherein the processing unit comprises one or more lock-in amplifiers for extracting the first sensor signal.

42. A system of insect sensors, comprising:

a plurality of optical insect sensor devices configured to be individually positioned within a geographic area, each of the plurality of optical insect sensor devices configured to:

monitor insect activity within a three-dimensional detection volume extending outside the insect sensor device by detecting light from the detection volume, wherein the detection volume is an enclosure-free void allowing unrestricted movement of living airborne insects into and out of the void, and

output detector data indicative of one or more optically detected attributes associated with respective detected insect detection events, each insect detec-

tion event being indicative of one or more insects being present in the detection volume.

43. A system according to claim **42**, further comprising a data processing system communicatively coupled to the plurality of optical insect sensor devices and configured to:

receive detector data from respective ones of the plurality of optical insect sensor devices, the detector data being indicative of one or more optically detected attributes associated with respective detected insect detection events, and

compute, from the received detector data, one or more quantities indicative of insect activity within the geographic area.

44. A system according to any one of claim **42**, wherein each insect sensor device comprises a detector module and a processing unit, the processing unit being configured to receive and process sensor signals from the detector module, and wherein the detector modules and/or the signal processing performed by the processing units of the respective insect sensor devices is calibrated according to a detector reference that is uniform across the plurality of insect sensor devices.

* * * * *