Delivering on a promise: Futureproofing automated insect monitoring methods

Roel van Klink[1,2,3*]

[1]German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig, Puschstrasse 4, 04103 Leipzig

[2]Department of Computer Science, Martin-Luther-University Halle-Wittenberg, Germany

[3]WBBS Foundation, Kanaaldijk 36, 9409 TV, Loon, The Netherlands.

* Corresponding author: roel.klink@idiv.de

ORCID: 0000-0002-8125-1463

Abstract

Due to rapid technological innovations, the automated monitoring of insect assemblages comes within reach. However, this continuous innovation endangers the methodological continuity needed for calculating reliable biodiversity trends in the future.

Maintaining methodological continuity over prolonged periods of time is not trivial, since technology improves, reference libraries grow, and both the hard- and software used now may no longer be available in the future. Moreover, because data on many species are collected at the same time, there will be no simple way of calibrating the outputs of old and new devices.

To ensure that reliable long-term biodiversity trends can be calculated using the collected data, I make four recommendations: (1) Construct devices to last decades, and have a five-year overlap period when devices are replaced. (2) Construct new devices to resemble the old ones, especially when some kind of attractant (e.g. light) is used. Keep extremely detailed metadata on collection, detection and identification methods, including attractants, to enable this. (3) Store the raw data (sounds, images, DNA extracts, radar/lidar detections) for future reprocessing with updated classification systems. (4) Enable forward and backward compatibility of the processed data, for example by in-silico data 'degradation' to match the older data quality.

Key words:

DNA barcoding, bioacoustics, computer vision, radar, lidar, monitoring, insects, arthropods, LTER

Main Text

The development of technological approaches for insect monitoring can allow unprecedented improvements in the spatial, temporal and taxonomic coverage of insect biodiversity assessments [1–4]. To meet the political, societal and industry needs for large-scale biomonitoring [5–7], these technologies can help close an important knowledge gap, since insects and other arthropods are the most species rich group of animals on earth, and perform important ecosystem services (e.g. crop pollination or decomposition) and disservices (e.g. disease transmission or crop damage). Insects are notoriously underrepresented in biodiversity monitoring schemes, since monitoring their diversity by traditional means with morphological identification is extremely time consuming and knowledge intensive. Moreover, some of the largest insect groups, such as flies and parasitoid wasps, are even within insect monitoring programs and ecological assessments rarely assessed. Automated monitoring could thus make large-scale insect biodiversity monitoring possible for a fraction of the effort and costs of traditional monitoring methods, and contribute to solving a number of identified challenges to large scale biomonitoring [6].

However, in order to reliably document changes in species occurrences, population sizes and biodiversity metrics over time, it is important to use the exact same method of monitoring over the whole sampling period. This applies to the collection, detection and identification methods, including any attractants used, as well as the taxonomic precision of the end product provided. This sounds logical, and even trivial, but anyone who has tried to do a sustained monitoring of biodiversity has learned that maintaining methodological continuity is not as easy as it sounds. Even when funding for continuous monitoring is secured (which is challenging even in the richest of countries), traps need to be replaced due to wear, loss or breakdown, workers learn to identify new species, fall ill or make mistakes, and taxonomy changes over time. In addition, there is a constant need for specialists with the right expertise, which is unfeasible in most parts of the world and for most taxa. For this reason, consideration of the methodology and data quality needed, is best done before monitoring commences.

Particularly when using high-tech devices and computer algorithms, the challenges to ensuring methodological continuity compound:

i. The hardware and software used in these devices are rapidly evolving and improving: camera sensitivity improves, barcoding pipelines change (Iwaszkiewicz-Eggebrecht this issue [8]), energy use becomes more efficient, etc. Although it is almost a moral imperative to use these developments to our advantage, and monitor as many species as possible for the lowest costs, we must also recognize the consequences of these developments for the long-term trends we're trying to calculate.
ii. Since the devices, which are often custom made for the purpose of insect monitoring, depend on hard- and software produced by third parties, there is no guarantee that these exact components will be available in the future. In fact, it is likely that they will not, because, industrial suppliers have no incentive to produce obsolete products, supply chains change, or new legislation may prevent the continued production or import of specific components.
iii. Weathering and wear of (parts of) the devices and traps in the field may make repeated use challenging, and parts may need to be exchanged regularly [see for example 9].
iv. The reference libraries of DNA barcodes, images and sounds used for classification are constantly growing, and will contain more and more species, allowing more accurate classification.
v. These devices are designed to collect multivariate data (dozens to thousands of species at the same time), and therefore, there will be no simple calibration possible of measured variables when monitoring devices are replaced with newer versions, especially given the volatility of insect population dynamics and the prevalence of rare species [10].

In most cases, technological improvements will increase detection and/or identification rates, which, when left unaccounted for, will lead to detecting a false increase in diversity over time. But any change in detection rates of any species will affect the inferences one can draw from the monitoring program in the future. The technologies covered in this Theme Issue (computer vision, DNA (meta)barcoding, radar and acoustics) are still in development, and are thus particularly vulnerable to the challenges outlined above. Although statistical methods may be able to account for some aspects of methodological variability, the reliability of the calculated temporal trends will suffer significantly from rapid methodological changes, in comparison to a continuous methodology..

I will illustrate the difficulties of ensuring methodological continuity over prolonged periods of time by two examples that are orders of magnitude less complex than any of the technologies discussed in this Theme Issue: Pitfall trapping of ground beetles (Coleoptera: Carabidae) with morphological species identification. In the north of the Netherlands, a program for monitoring ground beetle populations by means of standardized, year-round pitfall trapping was started in 1959 by the workers of the Willem Beijerink Biological Station, part of what is now Wageningen University. They started trapping ground beetles in custom-made square metal cans with an exact perimeter of 1m, [11,12]. These traps were replaced in the 1980's and possibly at an earlier time as well, but unfortunately this was not well documented. After the biological station was formally dissolved in 1998, the trapping program was continued by the volunteers of the WBBS foundation using the cans constructed in the 1980's. By 2020, the traps were in need of replacement, and we acquired funding for the construction of new traps.

Fig. 1. *The edges of the old (a) and the new (b) ground beetle traps. Due to technological changes, the old, rounded, edge would be excessively hard to reproduce. We have aimed to make the edge as similar as possible under field conditions (c). Photo's: Henk de Vries (a), Alje Woldering (b & c).*

Although we were unable to find back the company that constructed the original traps, this looked like a straightforward construction job to us, which any metalworking company could do. However, after numerous emails, phone calls and visits to various companies, we found that the technique for constructing the rounded edge of the old cans (Fig. 1a), a process called 'edge beading', had fallen out of use for this kind of sheet metal, and that a custom-made mold (a 'die') for a bead of exactly this size would be excessively expensive (roughly half of our budget for replacing the traps). We therefore had to settle for a different edge type for our new traps (Fig. 1b). We hope that, at least from a beetle's perspective, there will be no difference between the trap types (Fig. 1c). We have replaced the traps in two phases over 2022 and 2023 to test if and how the catch is affected by the trap replacement.

A second example from the same monitoring program is the challenge we have faced regarding the transition between data formats. All data collected on a weekly basis from 1959 to 1998 were once digitized, and stored on computer tapes. Currently, reading such tapes is close to impossible, especially since we don't know which computer brand was used for data entry, or the software format the data were stored in. Fortunately, all data are still available on paper sheets, and we are currently working on redigitising these, where we ensure compatibility with the upcoming Humboldt Extension for ecological inventories to the GBIF Darwin Core. That this is necessary illustrates the importance of a timely transition between data formats as hard- and software evolve. In 2009, Borer et al. [13] published some excellent advice on data management, and wrote: 'As hard as it is to believe today, we can foresee the day when CD-ROMs might be difficult to read.'. As per 2023, that day has come and gone, and it would be well advised to rapidly move all data stored on CD-roms and DVD's to the cloud (or better, to make them openly accessible on a FAIR biodiversity data portal like GBIF). This trend of soft- and hardware replacement is likely to continue, and it will be important to keep up with these developments.

Now imagine going through a similar process for replacing a modern camera trap, a radar, a sequencer or a barcoding pipeline, or to try to read data 20 years from now. Ideally, we would want every single hard- and software component used for detecting and identifying organisms, and for data storage to remain constant for as long as the monitoring lasts: several decades. But this is exceedingly unlikely, since all technological insect monitoring methods depend on a chain of industrial suppliers for the hard- and software used in the devices, as well as for data storage. These suppliers have no interest in continuing the production of obsolete products, just as we, as end users should use the best products available to monitor as many species as possible. Hence, we will need other solutions to ensure methodological continuity.

Below, I make four concrete recommendations, from the level of device construction to the processed biodiversity data, to ensure the data produced now can be used to calculate reliable biodiversity trends in the future. These recommendations are in most cases not only applicable to new technologies, but are equally useful for traditional insect monitoring programs:

a) Build to last. Design devices with the aim of lasting decades, and don't wait for them to break down before replacing them. Ideally, aim for an overlap period of 5 years when replacing devices, but here it should be considered that two traps set up in close proximity may influence each other, especially when an attractant is used. In such cases, a phased transition across multiple locations may be a better option.
b) Keep extremely detailed metadata, so that future devices can collect data in the same way, even when the sensors improve. This is especially important when an attractant, such as light or a colored screen is used, because a change in attractant(s) will inevitably affect insect behavior. But also extreme metadata detail is required regarding the sensitivity of the sensor(s), as this information can be used to make collected data more comparable. Metadata should thus include the exact light spectrum (including parts of the light spectrum that are not visible for humans, and luminosity of a light trap, exact screen color and texture [see 14 this issue], motion triggers (if used), camera resolution, microphone sensitivity, frequency range, and recording bitrate, sequencing depth, biochemical and bioinformatics pipelines for (meta)barcoding [see 8 this issue], etc. In addition, all data on the operational status of the traps and/or sensors, as well as the exact locations, should be recorded and stored. Although a lack of historic metadata may prevent us from precisely redoing historical investigations, we can make future resampling campaigns a lot more accurate.

c) Store all raw data (photos, condensed audio recordings, radar/lidar detections, barcoding libraries, etc.) in a non-proprietary format for future reprocessing using new algorithms, computational facilities and reference libraries. For this, a data infrastructure is needed that can handle and process the expected volume of raw data, and that can ensure data accessibility in the future. In addition, the energy, and thus environmental, costs of data storage and reprocessing should be considered.

d) Ensure forward and/or backward compatibility of the processed data (data with assigned taxonomic names), so that the quality of the data collected in the future can be made comparable to the data collected now, regarding, for example, the taxonomic depth and the sensor sensitivity. This may be done by either bringing currently collected data up to standards of the future (which will possibly need reprocessing, see previous point), or by in-silico degradation of future data to match the current standards (assuming that future data will be of higher quality than current data). To make this possible, there is a strong need for the automated taxonomic harmonization of species identifications. The GBIF taxonomic backbone, which is based the Catalogue of Life [15], the Barcode Index Numbers from the Barcode of Life project [16], and 103 other taxonomic resources [17], seems the most promising resource for automated harmonization with the most up-to-date taxonomic classification for both traditional and genetic data.

These recommendations do not only apply to the monitoring of insects, but to any type of automated biodiversity monitoring, for example camera trapping of mammals, acoustic monitoring of birds, bats, whales or fish, eDNA, or bird radar.

Conclusions

If the difficulties of securing long-term funding for biodiversity monitoring and the continued training of taxonomic specialists can be overcome, the technological developments of the past decades bring large-scale insect monitoring is closer than ever. But before we start deploying devices whenever an opportunity arises, it will pay off to first consider how we want to use these data now and in the future. What we can learn and infer, and for whom and for what purpose the data will be useful, will crucially depend on the choices we make today. For many purposes, including conservation planning and pest monitoring, accurate species level identifications are of crucial importance. Likewise, for calculating long-term trends, methodological continuity is crucial. If the above recommendations are followed, I am confident that automated insect monitoring will yield us many insights about the changes in insect biodiversity over the coming decades.

Acknowledgements:

This work was funded by DFG Grant FZT 118 to the German Centre for Integrative Biodiversity Research. The new ground beetle traps were funded by the Uyttenboogaart-Eliasen Stichting and the Prins-Bernhard Cultuurfonds. I thank Fons van der Plas and two reviewers for commenting on an earlier version of this manuscript, and Rikjan Vermeulen, Henk de Vries, Alje Woldering and Kees van der Laaken† for their dedication to sampling and identifying the beetles. I thank Henk de Vries and Alje Woldering also for the photos used in Fig 1.

References

1. Van Klink R *et al.* 2022 Emerging technologies revolutionise insect ecology and monitoring. *Trends Ecol. Evol.* **37**, 872–885. (doi:10.1016/j.tree.2022.06.001)

2. Gibb R, Browning E, Glover-Kapfer P, Jones KE. 2019 Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* **10**, 169–185. (doi:10.1111/2041-210X.13101)

3. Høye TT *et al.* 2021 Deep learning and computer vision will transform entomology. *Proc. Natl. Acad. Sci.* **118**, 2002545117. (doi:10.1073/PNAS.2002545117)

4. Bauer S *et al.* 2017 From agricultural benefits to aviation safety: Realizing the potential of continent-wide radar networks. *BioScience* **67**, 912–918. (doi:10.1093/biosci/bix074)

5. Gonzalez A *et al.* 2023 A global biodiversity observing system to unite monitoring and guide action. *Nat. Ecol. Evol.* , 1–5. (doi:10.1038/s41559-023-02171-0)

6. Moersberger H *et al.* 2022 Europa Biodiversity Observation Network: User and Policy Needs Assessment. , 218.

7. Potts SG *et al.* 2020 Proposal for an EU Pollinator Monitoring Scheme.

8. Iwaszkiewicz-Eggebrecht E, Zizka VMA, Lynggaard C. In press. placeholder. *Philos. Trans. R. Soc. B-Biol. Sci.*

9. Ssymank A, Sorg M, Doczkal D, Rulik B, Merkel-Walner G, Vischer-Leoplod M. 2018 Praktische Hinweise und Empfehlungen zur Anwendung von Malaisefallen für Insekten in der Biodiversitätserfassung und im Monitoring. *Ser. Nat.* **1**.

10. Fisher ARA, Corbet AS, Williams CB. 1943 The number of animals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

11. Den Boer PJ. 1990 Density limits and survival of local populations in 64 carabid species with different powers of dispersal. *J. Evol. Biol.* **3**, 19–48. (doi:10.1046/j.1420-9101.1990.3010019.x)

12. Den Boer PJ. 1977 *Dispersal power and survival - Carabids in a cultivated countryside*. Wageningen: Veenman en zonen B.V. See http://www.biological-station.com/Publicaties Piet den Boer2.html.

13. Borer ET, Seabloom EW, Jones MB, Schildhauer M. 2009 Some Simple Guidelines for Effective Data Management. *Bull. Ecol. Soc. Am.* **90**, 205–214. (doi:10.1890/0012-9623-90.2.205)

14. Roy DB, Hoye TT. In press. placeholder. *Philos. Trans. R. Soc. B-Biol. Sci.*

15. Bánki O *et al.* 2023 *Catalogue of Life Checklist*.

16. Ratnasingham S, Hebert PDN. 2007 bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes* **7**, 355–364. (doi:10.1111/j.1471-8286.2007.01678.x)

17. GBIF Secretariat. 2023 GBIF Backbone Taxonomy. Checklist dataset.