The evolito diversity index

Daniele Torelli
dato@evolito.earth

Páll Vang Kjærbo
pakj@evolito.earth

October 7, 2025

1 Introduction

Biodiversity, the variety of life forms on Earth, is essential for the functioning and stability of ecosystems[4, 8]. Understanding and quantifying biodiversity is therefore critical for conservation efforts and ecosystem management[10]. However, the sheer complexity of ecosystems often poses a formidable challenge in assessing biodiversity comprehensively, therefore the use of proxies and metrics becomes necessary[12, 14].

Among the myriads of organisms inhabiting Earth, insects are by far the most diverse group with over 1 million described species[23]. They therefore constitute approximately half of all $\approx$ 2.1 million described species in the world[9]. Terrestrial arthropod biomass constitutes around 200 Mt of which the majority are insects[5]. This is over three times the global biomass of humans and 20 times that of wild terrestrial mammals[1, 5]. In addition to being incredibly speciose and abundant, insects are found in virtually every terrestrial and aquatic ecosystem on the planet with almost every ecological life history imaginable[19].

Insects are therefore great candidates for monitoring biodiversity across the planet, but traditional methods of insect biodiversity monitoring are labor intensive and require biological expertise[13]. However, with the advent of automated sensor solutions such as the Volito[18] and evoSense sensor[15], gaining large-scale continuous monitoring of insect abundance and biomass is possible without intensive manual labor. These sensor estimates are based on registration of flying insects using photonic[18] and electrostatic field[15] technologies, which can detect and measure the insect's wing beat frequency (WBF).

In this paper we describe the principles of traditional biodiversity indices followed by the introduction of a novel index (evolito diversity index), based on the distribution of insect WBFs. The WBF data used is derived from the Volito sensor[18]. However, both the Volito and evoSense sensors[15] register WBFs of insects, so the approach outlined in this paper applies for both technologies. Since WBF to some degree is used to distinguish specific insect taxa, such as mosquito species[11], and WBF is, at least partly, explained by phylogenetic relationships[24], we propose that WBF distribution can function as a proxy for insect taxa distribution (evenness). We explore the rationale behind this approach, how it relates to traditional indices, its strengths and limitations, and the various ways in which it can contribute to our understanding of ecosystem dynamics, conservation efforts and human impact.

2 Background

2.1 Diversity indices[22]

The most important components of biodiversity are **species richness** S and **species evenness**. The first refers to the number of different species present in a community and it is a simple count of the number of unique species. A community with a high species richness has a greater variety of species, while a community with low species richness has fewer species.

The latter refers to the relative abundance of individuals among the different species in a community. It quantifies how evenly the individuals are distributed among the species. A community with high evenness has a relatively similar number of individuals of each species, while a community with low evenness is dominated by one or a few species, with others being rare. It is worth noting that measures of richness and evenness are not limited to species-level analyses but can also be applied to higher taxonomic levels such as family or genus. This is particularly relevant in studies of insect diversity, where identification to species may be challenging, yet lower taxonomic resolution can still yield meaningful ecological insights[2, 3, 25].

Various metrics and indices have been developed to quantify biodiversity concepts. The two prominent ones often cited in literature are the Simpson's diversity index and the Shannon-Wiener index.

Simpson's index[21]. The Simpson diversity index D, also known as the Gini-Simpson index, is a metric used in ecology to measure species diversity. It focuses on the probability that two individuals randomly selected from the community belong to different species. Thus, this index is particularly useful for identifying ecosystems with low diversity due to the dominance of a few species, which can be important in conservation efforts. If p_i is the proportion of the total individuals represented by species i (relative abundance), then it is calculated as

$$D = \sum_i p_i^2 \tag{1}$$

With only one species present, so in absence of diversity, the Gini-Simpson index is 1 and approaches to 0 as both species richness and evenness increase. Related metrics are the **Simpson's index of diversity** $1 - D$ and **Simpson's evenness** $(1/D)/S$, where S is the value of species richness.

Shannon's index[20]. The **Shannon diversity index**, also known as Shannon-Wiener index, is another widely used measurement in ecology to quantify the diversity of species within an ecosystem over time. It also takes into account both the species richness and their relative abundance. It can be obtained as

$$H_S = -\sum_i p_i \ln p_i \tag{2}$$

In absence of diversity the value is 0. It approaches the maximum value $H_{\text{MAX}} = \ln S$ when all S species are equally present. The **Shannon's evenness**, also known as Pielou's index[17] can then be calculated as

$$E_H = H_S / H_{\text{MAX}} \tag{3}$$

which is comprised between 0 and 1.

In terms of reflecting species evenness and richness, both indices have their strengths and weaknesses. Simpson's index may provide more insight into the dominance of particular species within a community. Because it squares the proportion of each species, it tends to give more weight to dominant species, potentially under-representing the influence of rare ones.

On the other hand, Shannon's index is less sensitive to the presence of dominant species and it is often seen as a better reflection of species evenness, as it gives equal weight to all species regardless of their abundance.

3 Methods

Our measure for biodiversity E_{ev} is based on the Shannon Entropy of the WBF distribution divergence from an homogeneous distribution. The WBF of an insect is the number of

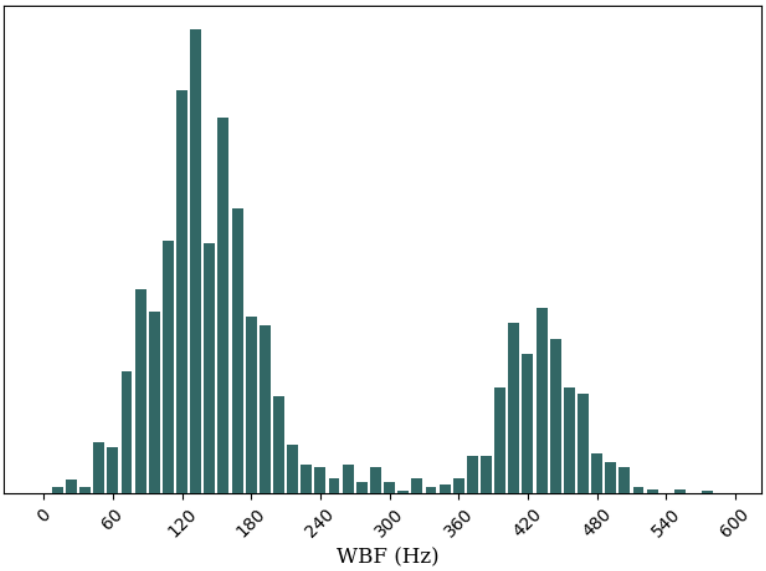


Figure 1: Example of a weekly WBF distribution for a field sensor in a specified interval of time.

times it flaps its wings in a second. It is the primary attributes calculated by analyzing the signal originated by such an event and recorded by our sensors. The Shannon Entropy[20] is a mathematical concept used to measure the diversity or uncertainty in a set of data and the formula to calculate is similar to the one in Eq. 2, but p_i represents the proportion of insect events with a WBF comprised within a certain interval (or bin) i

$$H_{ev} = \sum_i p_i \ln p_i \tag{4}$$

In Fig. 1 is shown an example of a WBF distribution for a specific session over a specific interval of time. Each bar represents the count of insects with a WBF within a certain bin. In the context of continuous distributions, the choice of bin size (or binning), so indirectly the choice of the number of bins N_{bin} within a certain range becomes crucial. The resulting number is then subtracted to the Shannon Entropy of a homogeneous distribution with the same binning, so $H_{ref} = ln(N_{bin})$ and

$$E_{ev} = 1 - (H_{ev} + H_{ref})/H_{ref} \tag{5}$$

It is worth noting that in Eq. 5, the maximum entropy reference value H_{ref} depends solely on N_{bin}, making it a constant determined a priori. In contrast, in Eq. 3, the maximum entropy reference H_{MAX} is a function of the richness S, and therefore varies depending on the sampled population.

Since H_{ev} is negative and in absolute value always lower or equal than H_{ref}, E_{ev} is a number comprised between 0 and 1. A higher value indicates less divergence from the homogeneous distribution, which is considered as the case with a great number of families (all bins are populated) with no dominant one (all families have the same abundance). When $H_{ev} \to -H_{ref}$ we have max evenness, in fact $E_{ev} \to 1$. On the other hand, when $H_{ev} \to 0^-$, which occurs for highly dominating families, then $E_{ev} \to 0$.

To assess and validate our method, labelling sessions, i.e. the recording of insect flight events in an enclosed setup containing only a specific family of insects, were carried out with representatives from 16 different insect families. The insect families used and the resulting WBF distributions are shown in Fig. 2. It can be noted that some families have irregular and overlapping distributions.

Using the data from the labelling sessions, we can control the exact number of insect families present for calculating the three diversity indices E_H, D and E_{ev}. This was done as follows: for each size of family richness S (ranging from 2 to 16), a randomized set of families

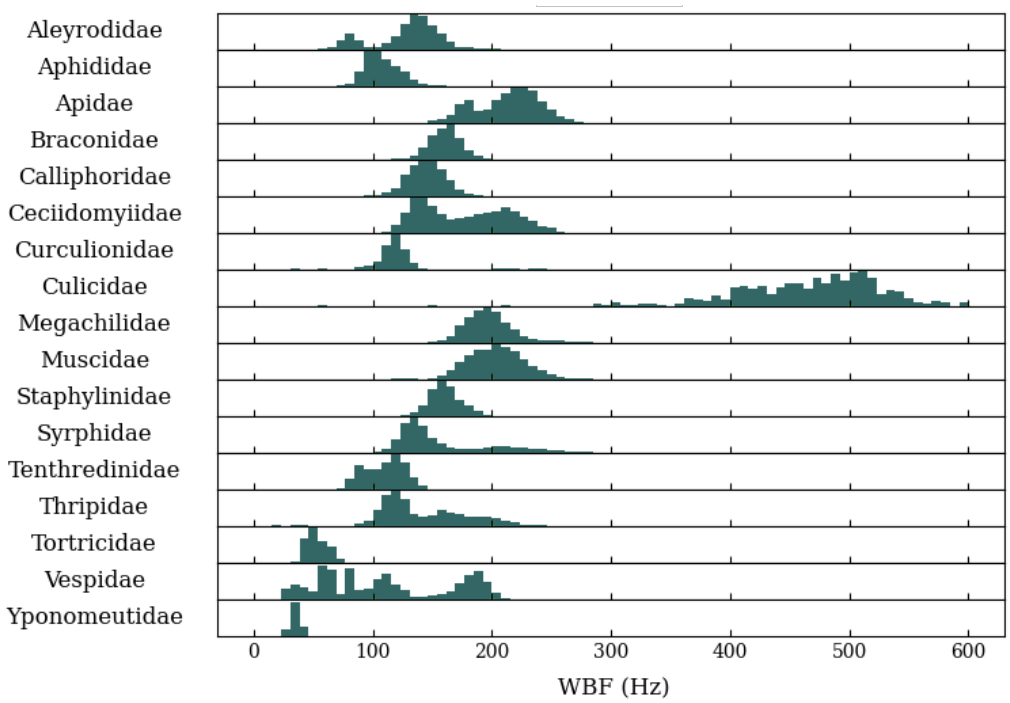


Figure 2: WBF distributions of labelled insect families.

has been sampled (e.g. [Apidae, Muscidae, Thripidae] when $S = 3$), as well as a randomized set of proportions p_i (e.g. [0.15, 0.55, 0.3]). Events from each family have been randomly sampled according to the proportions defined in p_i for a total of N_{TOT} events. These events collectively represent a session of events for which we draw a WBF distribution as the one in Fig. 1 and calculate evolito diversity index to compare to Shannon's evenness (which is bound between 0 and 1) and Simpson's index of diversity (which is expected to grow with S).
Note that p_i is enough for the calculation of the Shannon's evenness and Simpson's index of diversity, while for evolito diversity index we make use of the aggregated WBF distribution, formally ignoring the number of families present in the session. We repeat this process fifteen times for each size of family collections.

4 Results

In Fig. 3 the results for each process and for each method are shown as shadowed circles, while in solid lines are the averages for evolito diversity index, Simpson's diversity and Shannon's evenness. Parameters' choice ($N_{TOT} = 1000$, $N_{BIN} = 30$, $WBF_{range} = (25, 600)$) has been driven by numerical tests and heuristic evaluation, but further investigation may lead to improved results.

The average Pearson's correlation scores and p-values for a series of analysis are shown in Table 1. In Fig. 4a is better shown the strong correlation between Shannon evenness index and the evolito diversity index, after a linear fitting of data shown in Fig. 3. While the results obtained may suggest a very strong correlation, it's important to note that such a high correlation may not necessarily be observed in individual tests, due to statistical fluctuations and method limitations highlighted in the next section. In Fig. 4b is shown one of the many examples encountered during this analysis that none the less corroborate the correlation between our richness-agnostic method and the Shannon evenness index. We can see that for S = 7, the random selection of few sessions with strongly unbalanced proportions between families produces a drop in evenness. This affects the Shannon's evenness index (in yellow) more than the Simpson diversity index (in red) as expected, but the evolito diversity index (dark green) is able to capture this feature too.

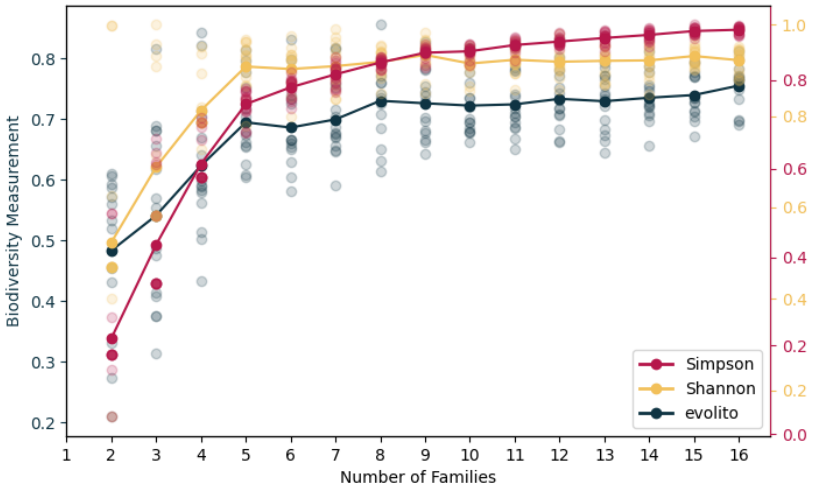
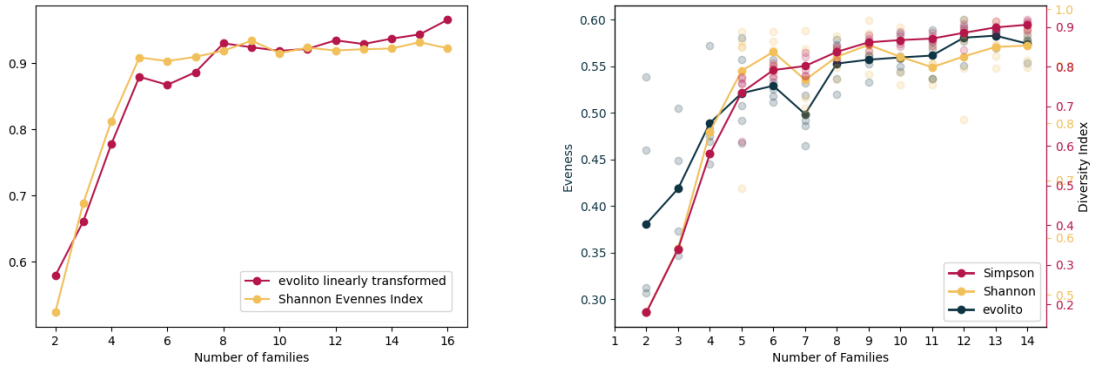


Figure 3: Statistical analysis and comparison between Shannon evenness and Simpson diversity.



(a) (b)

Figure 4: (a) linear transformation $Y = 1.42x - 0.11$ between Shannon evenness and evolito diversity index. (b) Case study in randomized sets.

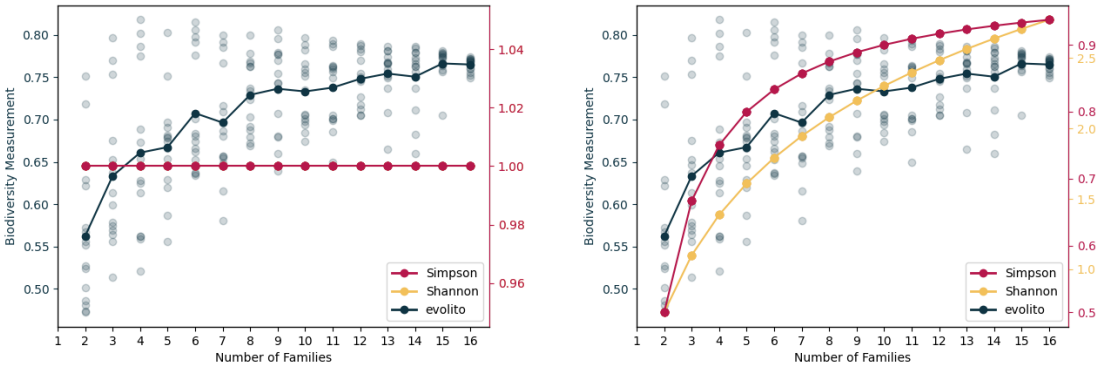
We finally conducted a series of stress tests to ascertain the predominant factors affecting our metric, whether it leans more towards insect family richness or evenness. The averages have been run over fifteen resampling of families (shadowed circles). Fig. 5 illustrates the scenario of maximum evenness, wherein all families are equally represented within the session, denoted by an array proportion of $[1/S] * S$. It is observed that the evolito diversity index exhibits an increasing trend with higher values of S. This trend suggests a notable influence of richness on the metric, albeit with a less pronounced steepness. Both the Shannon and Simpson evenness indices maintain a numerical value of 1 by definition (Figure 5a). Despite this, the evolito metric demonstrates a moderate correlation with the diversity indices, particularly with the reciprocal Simpson diversity function (which in this case becomes $D = 1 - (1/S)$) and also ranges between 0 and 1) (Figure 5b), rather than the logarithmic Shannon function ($H_S = logS$).

In Fig. 6, we examine the scenario of a session with a dominant family, where 70% of the insects within the sessions originate from a single family, while the remainder are randomly distributed among the other S-1 families. In this context, both the Shannon and Simpson evenness (see Fig. 6a) exhibit rapid declines as S increases, signifying the growing prominence of the dominant family (always set at 70%). Conversely, the diversity indices show

less susceptibility to dominance, with a gradual increase as S expands (in Fig. 6b). Notably, our evolito measurement appears to maintain an almost consistent, relatively low value, combining both the decline in evenness and the rise in richness as S increases.

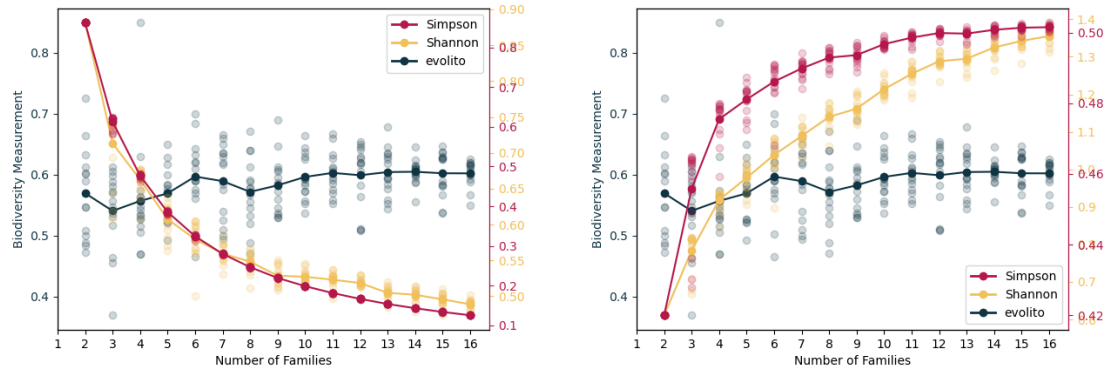
	Correlation	p-value
Shannon evenness	0.82	<0.05
Simpson evenness	0.62	<0.05
Simpson diversity	0.84	<0.05

Table 1: Pearson correlation for randomized sets.



(a) (b)

Figure 5: Maximum evenness test: (a) comparison with evenness indices and (b) comparison with diversity indices



(a) (b)

Figure 6: Dominant class test: (a) comparison with evenness indices and (b) comparison with diversity indices

5 Discussion and Conclusions

The numerical method employed to assess biodiversity within our study ecosystem, utilizing insect WBF distributions as a proxy for family-level biodiversity, has yielded valuable insights. However, it is essential to acknowledge its limitations to ensure the integrity of our findings.

- Incomplete Taxonomic Coverage. One notable limitation of our study is the incomplete taxonomic coverage of insect families. We acknowledge that not all families have been accounted for in our dataset, primarily due to taxonomic challenges and data availability constraints. As a result, our assessment may not provide a comprehensive representation of the ecosystem's true biodiversity used in this study.
- Data Quality and Reliability. It is known that insect behavior can be significantly influenced by environmental factors, including light[6], temperature and humidity[16, 7]. The data used for this analysis is from sensors in controlled environments, like cages. As a consequence, our analysis may not fully capture the nuances and variability of insect behavior (including their WBF) under field conditions. Moreover, in real-world sessions, rain or other forms of noise sources may lead to a number of false positive events with a WBF falling within the range of interest and altering the distribution (especially in the low frequency region): this is to a good extent already addressed in the data acquisition process by specifically developed event extractors procedures and rain/noise filters[15].
- Inherent complexities of WBF distributions. In Fig. 2 a significant overlap in WBF distributions can be observed across various insect families which can obscure the distinction between different families (e.g. Braconidae, Megachilidae, Muscidae,...). The method, reliant on Shannon entropy derived from these distributions, may fail to delineate the unique contributions of individual insect species to overall biodiversity. Further research into examining features other than the fundamental WBF (such as WBF harmonics), could prove useful in enhancing differentiation between insect families. Furthermore, the presence of certain insect groups, such as Culicidae (mosquitoes), with disproportionately high wing beat frequencies, exacerbates this issue. Especially for real-world field sessions, they may contribute with a pronounced peak at higher frequencies, distorting the overall shape and skewness of the distribution. For the same reason, they may yield to a significant gap within the range of 300 and 400 Hz, contributing to the sparsity of the distribution and complicating the numerical calculation of evenness.

Despite these limitations, our numerical method has yielded promising results and these findings align well with other widely-used biodiversity metrics in the literature. The current consistency underscores the validity of our method and suggests that it can provide valuable insights into the ecosystem's ecological health. Future research should prioritize validating our numerical method against real-world sensor data. Comparing these sensor-derived insights with established benchmarks such as Malaise trap diversity data could enhance the robustness of our approach or the development of new ones. This alignment with empirical data would not only strengthen the method's applicability in practical settings but also highlight its potential to complement or even improve upon traditional biodiversity assessment tools.

References

[1] Y. M. Bar-On, R. Phillips, and R. Milo. The biomass distribution on earth. Proceedings of the National Academy of Sciences of the United States of America, 115:6506–6511, 6 2018.

[2] 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 and Applied Ecology, 4:589–593, 2003.

[3] M. M. Driessen and J. B. Kirkpatrick. Higher taxa can be effective surrogates for species-level data in detecting changes in invertebrate assemblage structure due to disturbance: a case study using a broad range of orders. Austral Entomology, 58:361–369, 5 2019.

[4] S. Díaz, J. Fargione, F. S. C. C. III, and D. Tilman. Biodiversity loss threatens human well-being. PLoS Biol, 4:e277, 2006.

[5] P. Eggleton. The state of the world's insects. Annual Review of Environment and Resources, 45:61–82, 2020.

[6] S. T. Fabian, Y. Sondhi, P. E. Allen, J. C. Theobald, and H. T. Lin. Why flying insects gather at artificial light. Nature Communications, 15, 12 2024.

[7] E. G. Farnworth. Effects of ambient temperature, humidity, and age on wing-beat frequency of periplaneta species. J. Insect Physiol, 18:827–839, 1972.

[8] IPBES. Global assessment report of the intergovernmental science-policy platform on biodiversity and ecosystem services. IPBES secretariat, 2019.

[9] IUCN. IUCN Red List (2025) "Table 1a: Number of species evaluated in relation to the overall number of described species, and numbers of threatened species by major groups of organisms". https://www.iucnredlist.org/resources/summary-statistics#Summary%20Tables, 2025. Version 2025-1, last updated: 27/03/2025. Accessed: 2025-05-07.

[10] J. P. Jones, B. Collen, G. Atkinson, P. W. Baxter, P. Bubb, J. B. Illian, T. E. Katzner, A. Keane, J. Loh, E. Mcdonald-Madden, E. Nicholson, H. M. Pereira, H. P. Possingham, A. S. Pullin, A. S. Rodrigues, V. Ruiz-Gutierrez, M. Sommerville, and E. J. Milner-Gulland. The why, what, and how of global biodiversity indicators beyond the 2010 target. Conservation Biology, 25:450–457, 6 2011.

[11] D. Kim, T. J. DeBriere, S. Cherukumalli, G. S. White, and N. D. Burkett-Cadena. Infrared light sensors permit rapid recording of wingbeat frequency and bioacoustic species identification of mosquitoes. Scientific Reports, 11, 12 2021.

[12] C. Loehle. Challenges of ecological complexity. Ecological Complexity, 1:3–6, 3 2004.

[13] G. A. Montgomery, M. W. Belitz, R. P. Guralnick, and M. W. Tingley. Standards and best practices for monitoring and benchmarking insects. Frontiers in Ecology and Evolution, 8, 1 2021.

[14] E. K. Morris, T. Caruso, F. Buscot, M. Fischer, C. Hancock, T. S. Maier, T. Meiners, C. Müller, E. Obermaier, D. Prati, S. A. Socher, I. Sonnemann, N. Wäschke, T. Wubet, S. Wurst, and M. C. Rillig. Choosing and using diversity indices: Insights for ecological applications from the german biodiversity exploratories. Ecology and Evolution, 4:3514–3524, 9 2014.

[15] F. B. Odgaard, P. V. Kjærbo, A. H. Poorjam, K. Hechmi, R. M. Luciano, and N. Krebs. Automated insect detection and biomass monitoring via AI and electrical field sensor technology. Scientific Reports, 15(1):29858, 2025.

[16] A. R. Parmezan, V. M. Souza, I. Žliobaitė, and G. E. Batista. Changes in the wing-beat frequency of bees and wasps depending on environmental conditions: a study with optical sensors. Apidologie, 52:731–748, 8 2021.

[17] E. C. Pielou. The measurement of diversity in different types of biological collections. J. Theoret. Biol, 13:131–144, 1966.

[18] K. Rydhmer, E. Bick, L. Still, A. Strand, R. Luciano, S. Helmreich, B. D. Beck, C. Grønne, L. Malmros, K. Poulsen, F. Elbæk, M. Brydegaard, J. Lemmich, and T. Nikolajsen. Automating insect monitoring using unsupervised near-infrared sensors. Scientific Reports, 12, 12 2022.

[19] T. Schowalter. Chapter 1 - overview. In Insect Ecology, pages 1–24. Academic Press, 5th edition, 2022.

[20] C. E. Shannon. A mathematical theory of communication. The Bell System Technical Journal, 27:379–423, 1948.

[21] E. Simpson. Measurement of diversity. Nature, 163, 4 1949.

[22] T. Smith and R. Smith. Elements of Ecology. Always learning. Pearson, 2015.

[23] N. E. Stork. How many species of insects and other terrestrial arthropods are there on earth? Annual Review of Entomology, 63:31–45, 2018.

[24] M. Tercel, F. Veronesi, and T. Pope. Phylogenetic clustering of wingbeat frequency and flight-associated morphometrics across insect orders: Flight strategies between insect orders. Physiological Entomology, 43, 03 2018.

[25] Y. Zou, W. van der Werf, Y. Liu, and J. C. Axmacher. Predictability of species diversity by family diversity across global terrestrial animal taxa. Global Ecology and Biogeography, 29:629–644, 4 2020.