Novel community data in ecology-properties and prospects

Florian Hartig[1,*], Nerea Abrego[2], Alex Bush[3], Jonathan M. Chase[4], Gurutzeta Guillera-Arroita[5], Mathew A. Leibold[6], Otso Ovaskainen[2,7], Loïc Pellissier[8,9], Maximilian Pichler[1], Giovanni Poggiato[10], Laura Pollock[11], Sara Si-Moussi[10], Wilfried Thuiller[10], Duarte S. Viana[12], David Warton[13], Damaris Zurell[14], Douglas W. Yu[15, 16]

1) Theoretical Ecology, University of Regensburg, Regensburg, Germany
2) Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35 (Survontie 9C), FI-40014 Jyväskylä, Finland
3) Lancaster Environment Centre, Lancaster University, UK
4) German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
5) Pyrenean Institute of Ecology, CSIC, Spain
6) Biology Department, University of Florida, Gainesville, Florida, USA
7) Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, P.O. Box 65, Helsinki 00014, Finland
8) Ecosystems and Landscape Evolution, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, 8092 Zurich, Switzerland
9) Unit of Land Change Science, Swiss Federal Research Institute for Forest, Snow and Landscape Research (WSL), 8903 Birmensdorf, Switzerland
10) Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, F38000, Grenoble, France
11) Department of Biology, McGill University, Montreal, Quebec, Canada
12) Doñana Biological Station (EBD-CSIC), Spain
13) University of New South Wales, Australia
14) University of Potsdam, Potsdam, Germany
15) Kunming Institute of Zoology; Yunnan, China
16) University of East Anglia, Norfolk, United Kingdom

* corresponding author, florian.hartig@ur.de

Keywords: environmental DNA (eDNA), community ecology, macroecology, passive acoustic monitoring, camera traps, joint Species Distribution Models (jSDMs)

Abstract

New technologies for biodiversity monitoring such as eDNA, passive acoustic monitoring or optical sensors promise to generate automated spatio-temporal community observations at unprecedented scales and resolutions. Here, we introduce "novel community data" as an umbrella term for these data. We review the emerging field around novel community data, focusing on new ecological questions that could be addressed; the analytical tools available or needed to make best use of these data; and the potential implications of these developments for policy and conservation. We conclude that novel community data offer many opportunities to advance our understanding of fundamental ecological processes, including community assembly, biotic interactions, micro- and macroevolution, and overall ecosystem functioning.

Novel community data – introduction and definition

Understanding the factors that govern the distribution of Earth's biodiversity across space and time remains one of the most pressing problems in biodiversity science. While human activities are rapidly altering the structure of biodiversity and the services it provides to humans [1], our ability to describe, model, and manage these

changes is hampered by the fact that conventional **biodiversity monitoring** is limited in spatial, temporal, and taxonomic scale and resolution, and is often poorly standardized and structured [2].

In recent years, major technological innovations in sensor technologies have occurred that promise to automate biodiversity monitoring. These include **environmental DNA** (eDNA, see Glossary), **passive acoustic monitoring** [3–5] and **visual sensors** (e.g. camera traps, see [6]), which, coupled with appropriate machine learning or deep learning **pipelines** [7,8], are moving the field "towards the fully automated monitoring of ecological communities" [9,10]. Hereafter, we refer to such **community inventories** generated by automated sensors and pipelines that do not directly involve humans in species detection and identification as **novel community data** (see also [11]).

The emergence of novel community data is likely to transform the way species distribution and abundance data are generated for the rest of the 21st century (e.g. [12–14]). The efficiency gains are such that hundreds or even thousands of species can be routinely detected and potentially quantified in their abundance across entire landscapes, resulting in a 'many-row, many-column' **community matrix**. These datasets are larger and richer in information than traditional community inventories, but they also have complicated properties such as higher rates of false positives or, for eDNA, unreliable information on relative abundance between species [15,16]. Novel community data therefore require appropriate statistical tools that can exploit

their higher information content while also accounting for their added complications [17].

The sensors and technologies used to generate novel community data have been extensively reviewed elsewhere [9,11,12,18–24]. In this review, we will therefore only briefly cover this topic and instead focus on how the combination of novel community data with new statistical tools both compels and enables us to transform data analysis, expand our scientific reach, and improve biodiversity conservation and management.

What makes novel community data really novel?

Over the past two decades, ecologists have assembled large collections of spatial occurrence or abundance observations (e.g. GBIF, IUCN range maps, or taxa-specific monitoring schemes). These data are frequently used in **species distribution models** (SDMs, e.g. [25,26]) to estimate species' environmental niches, project future distributions under climate or land-use change, or generate biodiversity metrics for conservation and management. A commonly recognized limitation of these data, especially when they are opportunistically collected, is uncertainty about observation errors and intensities [27]. Moreover, these data are rarely suitable for inferring local community co-occurrences across trophic groups, limiting their potential for understanding the role of **biotic interactions** in community and ecosystem dynamics.

Dedicated conventional data collection schemes exist that provide both presence and (somewhat reliable) absence or abundance/biomass information for entire local communities across space [28]. However, using conventional survey techniques, such data are typically limited in sample size, spatial and temporal extent and especially in taxonomic coverage and resolution (see [20], but see [29]).

The emergence of novel community data (Fig. 1) promises to fundamentally alter this established landscape of biodiversity observations. It is tempting to dismiss our ability to sequence environmental DNA (eDNA), ancient DNA (aDNA), and bulk-sample DNA [20,21,24,30], as well as the availability of camera traps or passive acoustic monitoring as merely a convenient way to generate more data (i.e. big data) of the same kind that we have been collecting. Such a view, however, neglects the many other dimensions on which novel community data differ from traditional community inventories.

Structure and standardization

Especially as technology evolves and pipelines are shared, compared, and converge on common standards, novel community datasets have the potential to be more structured and standardized than traditional sampling schemes. Moreover, novel community data are typically generated according to a fixed plan using low-expertise collection methods, positive and negative controls, and a standardized processing pipeline for species identification. Therefore, results are less dependent on individual observers.

Importantly, by standardized, we do not mean error-free. eDNA data, for example, can have considerable errors (see Box 1). However, because these errors are usually more consistent and therefore somewhat predictable, they can be more easily corrected using statistical methods than errors in conventional surveys that arise from different human observers or subtle differences in sampling protocols.

Spatial, temporal, and taxonomic extent and resolution

A second difference is that the automated way in which novel community data are generated makes them scalable to high spatial, temporal, and taxonomic resolution [30,31]. Different sensors have different strengths along these dimensions, but those can be combined by using multiple sensor types (see also [14]).

For example, while eDNA data have particular strengths in taxonomic breadth and resolution, as well as detection sensitivity and hence community completeness (Box 1), acoustic and visual sensors are better at producing continuous community time series. Indeed, acoustic and visual sensors offer the unique opportunity to continuously capture biodiversity over daily, seasonal, and even decadal time scales, something difficult to achieve with non-automated sampling schemes. An obvious advance for the field would be to use statistical methods to combine observations from these different data streams into a combined spatiotemporal data product or model (cf. [20,32], see also outstanding questions).

All sensors can in principle also be used to estimate abundance, although this will typically require additional steps (for eDNA, see [33] and Box 1). Next-generation

methods may even allow individual-level identification and tracking (via genetic data or image analysis) to investigate behavior, dispersal or migration patterns. Moreover, with eDNA, we can also identify taxonomic patterns below and beyond the species level, for example **exact-sequence variants (ESVs)** or genetic diversity within and between species [34].

Metadata acquisition and matching to other data sources

Another advantage of using standardized sensors, rather than humans, is that **metadata** can be easily recorded during data acquisition. Metadata typically includes time and location stamps, but importantly, also instrument errors and taxonomic uncertainties, which are rarely recorded in conventional surveys. Universally available metadata on time, location and taxonomy facilitates matching observations to other local sensors and independent data products, such as weather stations, remote sensing data, phylogenetic or trait information, or biotic interactions extracted from visual, acoustic data, or eDNA analysis [35]. The resulting combined data products could be of interest as essential biodiversity variables for the GEO-BON platform [36]. We acknowledge that the collection of rich metadata is best practice for conventional biodiversity inventories as well; however, we believe that in practice, automated sensors are likely to collect richer and more structured metadata than conventional surveys.

Observation errors and data quality

Despite these advantages, ecologists are often skeptical about the quality and reliability of novel community data. We recognize that each sensor type presents certain technical challenges, some inherent in the measurement process (e.g. the field of view of a camera) and others in the analysis pipeline (e.g. for eDNA, incomplete **DNA barcoding** reference databases or PCR errors; or for acoustic and visual sensors, transferability of deep learning methods for species recognition). The two-step process of the measurement itself and the pipeline for analysis and species identification can introduce errors and biases that are more complex than for conventional data (see Box 1 for a discussion of the eDNA pipeline). However, the development towards standardized pipelines and protocols, as well as the collection of rich metadata, also offers many opportunities to account for such errors in subsequent statistical analyses (see later section “Statistical models to deal with observation errors”).

Using novel community data to answer long-standing ecological questions

Having established that novel community data will provide not only a larger sample size, but also a richer, more standardized, and more interconnected data product than traditional biodiversity monitoring data, we focus on how these data will transform the way we can approach classical and new ecological questions. We

organize this discussion around five themes: i) **species associations**, ii) biotic, especially trophic interactions, iii) beyond the species concept, iv) real-time monitoring and long time series, v) understanding ecosystems as complex systems.

Species associations

Because novel community data can provide complete community inventories, they are well suited for investigating species associations. Raw species associations can arise from shared environmental preferences, but even when this is accounted for (see section “Statistical tools”), species often show remaining associations. These associations may be artifacts due to unmeasured or inadequately measured environmental or spatial factors [e.g. 37–39], but they may also reflect biotic interactions. The ability to comprehensively quantify species associations, especially when used in conjunction with direct observations of biotic interactions (see next subsection), offers the potential to advance the long-standing goal of disentangling spatial, abiotic and biotic factors as drivers of (meta)community assembly [40–42]. Moreover, if the data contain both spatial and temporal dimensions, associations can be investigated over both time and space, which may be critical to infer the underlying processes of metacommunity assembly [43]. Finally, even if the causes of spatial associations cannot be resolved, they reduce unexplained variation in the community composition and thus may provide a more realistic estimate of the irreducible stochasticity in community dynamics and assembly rules (e.g. [41]).

Biotic interactions

Novel community data, particularly eDNA data, can also be used to directly infer species interactions, both trophic and mutualistic [44]. The most straightforward way to observe trophic interactions and thus infer entire food webs is to sequence the gut contents of individuals (see, for example, [45], who sequenced the gut contents of coral reef fish to reconstruct a complex marine food web). It is also possible to infer host-vector-pathogen networks [46] or mutualistic interaction networks from interaction residues, e.g. by analyzing pollen on pollinators [47] or eDNA traces on flowers [35]. Such direct observations of species interactions can be compared to species associations or disturbances data (e.g. [48]) to understand how biotic interactions affect community assembly, ecosystem dynamics or species distributions.

Beyond the species concept

Another area where in particular eDNA data could lead to advances is in challenging the near-exclusive role of species as the basic unit for quantifying biodiversity and community patterns. While we believe that the species concept will remain central to ecology, novel community data can increase taxonomic resolution to the subspecies or even ESV level. This would not only solve the problem of cryptic species [49] but could also reveal large-scale ‘macrogenetic’ patterns of interspecific genetic variation and gene flow (cf. [50,51]). An important question is how such a more “granular” view of a species’ distribution could be integrated into concepts such as

competition, distribution, the niche, or extinctions, which are central to both ecology and practical conservation (e.g. [52,53]).

Real-time monitoring, nowcasting and ancient DNA

A natural advantage of acoustic and visual sensors over eDNA is their high temporal resolution, which offers the potential to observe short-term changes in population size, species interactions or habitat preferences, or phenological changes as well as community time series (e.g., [21], Fig. 2). Together, this offers the potential for real-time monitoring and nowcasting of biodiversity changes, biological invasions and pathogen outbreaks [54,55]. Another interesting idea is the ability to generate observations and time series from the past using ancient DNA [21,56], which could be critical for understanding human impacts on ecosystems in the Anthropocene.

Ecosystems as complex systems

Finally, the fact that novel community data provide direct measurements of species interactions (i.e. the trophic structure) together with community inventories at high spatiotemporal resolution may help us to revive the old aspiration of “modelling all life on Earth” [57], i.e. understanding ecosystems holistically as complex systems and describing their various interactions through mechanistic ecosystem or macroevolutionary models (e.g. [58]).

Statistical tools for novel community data

The “law of the instrument” famously warns us that “if all you have is a hammer, everything looks like a nail”. The saying cautions us that instruments and analytical tools, rather than scientific curiosity, often determine what research questions are asked. While the availability of new sensors expands our toolbox for data collection, tailored analytical approaches for novel community data are still rare, which currently limits our ability to use these data for answering the ecological questions we listed in the previous section. We see three main directions in which statistical methods for novel community data should be developed: community and metacommunity analysis, time series analysis, and network analysis.

Community and metacommunity analysis

Community and metacommunity analysis aims to understand how community composition changes as a function of the environment and possibly interactions between communities. Statistically, we can approach this problem from at least three angles: we can use differences or changes in community composition as a response (e.g. ordination, Mantel tests or regressions on distance matrices [59]); we can use constrained ordinations to partition effects on community composition between spatial and environmental predictors; or we can develop statistical models that predict community composition directly (as done, for example, in joint species distribution models (jSDMs), see [60–62] and Box 2). While each of these approaches has its strengths, we find the option of modelling communities directly

253 with jSDMs particularly promising because it allows us to infer species-specific
254 environmental preferences, spatial effects, and species associations, all of which are
255 quantities that are biologically interpretable and useful for making predictions.

256 Time series to infer causal drivers

257 Apart from a few exceptions, conventional monitoring has been unable so far to
258 provide continuous time series over large spatial scale and long periods of time. This
259 is unfortunate, because time series are better suited than static data for separating
260 correlation from causation. A prominent idea in causal time series analysis is the
261 concept of Granger causality [63], which posits that because the cause must
262 precede the effect, we can regress our observations (in this case the community
263 composition at each time step) against the observations of previous time steps. This
264 approach could also be used to infer asymmetric interactions (and thereby
265 hierarchical competition), and it has been argued that interactions based on such a
266 temporal or spatio-temporal approach are more likely to match with true biotic
267 interactions (see [64] and Fig. 2, for an implementation in an extended jSDM). Novel
268 community data, and especially acoustic and visual sensors, can provide continuous
269 time series data at unprecedented rates. Therefore, we believe that these data could
270 be instrumental in inferring causal relationships between species or groups of
271 species and in better understanding community assembly as a whole.

Network analysis

A third avenue for statistical analysis is to analyze and compare species association networks inferred through jSDMs and networks of mutualistic, trophic or competitive biotic interaction networks that are generated, for example, by sequencing gut contents (see also Fig. 1). This line of research could leverage methods from the field of network analysis [65], which often struggles with the same data limitations as in community ecology. Novel community data could allow us to analyze larger and more complex networks (e.g. [66]), analyze how these networks change across environmental gradients [67], and link these patterns to community data to understand how biotic interactions, in conjunction with environment and space, give rise to spatio-temporal biodiversity patterns [68]. For example, it has been found that species associations change with scale [69], but it is unclear whether such changes reflect anything about their underlying biotic interactions. Another example is that although two species interact locally (e.g. predator-prey), they may not show any association [70]. Understanding the interplay between association and interaction networks may be key to understanding the role of biotic interactions in structuring communities and spatial biodiversity patterns.

Statistical models to deal with observation errors

When designing these and other statistical analyses for novel community data, it will likely be critical to incorporate observation models that account for detection probabilities and taxonomic uncertainties. Observation models are not specific to

293 novel community data, but detection errors may be more pronounced and
294 complicated in novel community data (e.g. Box 1). On the positive side, due to
295 standardized pipelines and rich metadata, errors and uncertainties in detection and
296 taxonomic assignment may be easier to estimate. Currently, statistical models are
297 emerging that correct species detections for false positives and negatives (e.g.
298 [71,72]) and that extend these ideas to communities and jSDMs [73,74], relative
299 biomass estimates [75] and continuous-score observations [76]. A challenge for the
300 future is to make these models more broadly accessible and ready for the
301 computational demands of large novel community datasets.

302 Improving predictions of biodiversity responses to global
303 change

304 Finally, novel community data could help to improve predictions of biodiversity
305 dynamics under global or climate change beyond the trivial fact that more data is
306 always useful. For example, spatio-temporal community data are better suited to
307 identify causal effects and directional interactions ([63], see also section “Time
308 series and causality”). Identifying these factors is particularly important when
309 predicting species or biodiversity responses outside present climatic conditions.

Leveraging novel community data to achieve socio-ecological resilience

Beyond scientific progress, novel community data may also enhance society's ability to create effective governance of biodiversity as a public good. In their seminal paper, Dietz *et al*. [77] describe five elements for the successful governance of public goods: (1) knowledge generation, (2) infrastructure provision, (3) political bargaining, (4) enforcement, and (5) institutional redesign.

The most obvious role for novel community data is to contribute to the first element: the generation of *high-quality*, *granular*, and *timely* information on ecosystem status, health and change, uncertainty levels, values, and the magnitude and direction of anthropogenic impacts. In addition, as new *infrastructure* allows methods to become more automated, independent parties can collect, analyze and compare large biodiversity datasets, making this knowledge more *understandable* and *trustworthy* [78]. Information with these properties can in turn make *political bargains* more achieveable and enforcement more effective. Governments can apply ‘technology forcing’ to encourage the creation of novel community data [79] and, ultimately, *redesign environmental institutions* for greater effectiveness, as exemplified by the UK’s Great Crested Newt offset market (Box 3).

Moreover, novel community data could also provide opportunities to redesign scientific and political structures. For instance, although most regulatory uses of eDNA still involve only single-species detection [79], in the US, these data are being

combined into a multi-species database, the Aquatic eDNAtlas Project. To facilitate such a process, rigorous sampling protocols, reference datasets and pipelines for creating biodiversity data (e.g. AI models for species recognition, barcode databases) should be applied that are freely available and integrated into global monitoring schemes and databases such as GBIF, IUCN, and GEOBON (e.g. [22,80]). Based on these, policy-relevant data products such as global biodiversity integrity maps with granular and timely data (e.g. STAR, see [81]) could be created. Bayesian optimal design could be used to identify data gaps and thus to prioritize funding for initiatives to fill these gaps. For industry, the availability of such data can help to integrate ecological impacts into corporate decision making. For example, the Task Force on Nature-Related Financial Disclosures (TNFD, tnfd.global) has developed an analytical framework for assessing corporate exposure to nature-related risks and opportunities.

Concluding remarks: Outlook for ecological research

Novel community data offer exciting opportunities for understanding and predicting biodiversity patterns. For the first time, we can hope to generate spatiotemporal community inventories with high spatial, temporal, and taxonomic resolution, in conjunction with traits, abiotic predictors, and observed true biotic (mutualistic and trophic) interactions. While the need for and value of such multi-faceted biodiversity data has been acknowledged for some time, the emergence of sensors that

inherently produce community rather than single-species data at scale have brought the achievement of this long-held goal within our immediate reach.

The lower cost, more complex structure, and the higher information density of these data have important implications for how we can conduct and advance ecological analysis, concepts and theories. We have argued that (joint) species distribution models, network analysis and time series, paired with statistical tools inherited from causal analysis, could serve as some of the core analytical tools to connect these data to important ecological research questions, particularly in niche theory, metacommunity theory, and network theory. Beyond this, novel community data also have high potential to provide crucial information for environmental management and biodiversity conservation.

Challenges for the future (see “Outstanding questions”) include the creation of appropriate data products, which includes establishing standardized field designs and pipelines and bringing together existing data in common databases, the establishment of accessible statistical models to analyze these data, and the use of these analytical tools to produce ecological theory as well as actionable predictions for management and conservation.

Acknowledgements

This paper originated from several meetings of the sCom working group supported by sDiv, the Synthesis Centre of iDiv (DFG FZT 118, 202548816). We thank the editor and two anonymous reviewers for their helpful comments on earlier drafts of

this manuscript, and Andrew Buxton and Sarah Garratt from NatureSpace Partnership for comments on the newt offset market. WT and GP also acknowledge support from the European Union's Horizon Europe under the project NaturaConnect (grant agreement No 101060429) and OBSGESSION, as well as the French Agence Nationale de la Recherche (ANR) through the Gambas project (ANR-18-CE02-0025). SS-M and WT were funded by the French Biodiversity Office through the FrenchBiodiv project. OO was funded by Academy of Finland (grant no. 336212 and 345110) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 856506; ERC-synergy project LIFEPLAN). GGA is currently supported by a ´Ramón y Cajal´ grant (RYC2020-028826-I) funded by the Spanish Ministry of Science and Innovation, the Agencia Estatal de Investigación (10.13039/501100011033) and "ESF Investing in your future". MAL was funded by U.D. NSF award 2025118. DSV is currently supported by a ´Juan de la Cierva Incorporación´ postdoctoral grant (IJC2020-044545-I) funded by the Spanish Ministry of Science and Innovation. DWY was supported by the CAS Strategic Priority Research Program (XDA20050202) and the CAS Key Research Program of Frontier Sciences (QYZDY-SSW-SMC024).

References

1. Díaz, S. *et al.* (2019) Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* 366, eaax3100

2. Pollock, L.J. *et al.* (2020) Protecting biodiversity (in all its complexity): new models and methods. *Trends Ecol. Evol.* 35, 1119–1128

3. Gibb, R. *et al.* (2019) Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* 10, 169–185

4. Sugai, L.S.M. *et al.* (2019) Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience* 69, 15–25

5. Darras, K. *et al.* (2019) Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. *Ecol. Appl.* 29, e01954

6. Tabak, M.A. *et al.* (2019) Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods Ecol. Evol.* 10, 585–590

7. Tuia, D. *et al.* (2022) Perspectives in machine learning for wildlife conservation. *Nat. Commun.* 13, 1–15

8. Pichler, M. and Hartig, F. (2023) Machine learning and deep learning—A review for ecologists. *Methods Ecol. Evol.* 14, 994–1016

9. Besson, M. *et al.* (2022) Towards the fully automated monitoring of ecological communities. *Ecol. Lett.* 25, 2753–2775

10. Bohan, D.A. *et al.* (2017) Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends Ecol. Evol.* 32, 477–487

11. Tosa, M.I. *et al.* (2021) The rapid rise of next-generation natural history. *Front. Ecol. Evol.* 9, 698131

12. Van Klink, R. *et al.* (2022) Emerging technologies revolutionise insect ecology and monitoring. *Trends Ecol. Evol.* 37, 872–885

13. Lin, M. *et al.* (2021) Landscape analyses using eDNA metabarcoding and Earth observation predict community biodiversity in California. *Ecol. Appl.* 31, e02379

14. Wägele, J.W. *et al.* (2022) Towards a multisensor station for automated biodiversity monitoring. *Basic Appl. Ecol.* 59, 105–138

15. Cristescu, M.E. and Hebert, P.D.N. (2018) Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. *Annu. Rev. Ecol. Evol. Syst.* 49, 209–230

16. McLaren, M.R. *et al.* (2019) Consistent and correctable bias in metagenomic sequencing experiments. *Elife* 8, 46923

17. Yu, D.W. and Matechou, E. (2021) The contribution of DNA-based methods to achieving socio-ecological resilience. In: Understanding ecosystems and

resilience using DNA, pp. 145-200. Science Report 485 SC190006/R, Environment Agency, Bristol. https://www.gov.uk/government/publications/understanding-ecosystems-and-resilience-using-dna

18. Wäldchen, J. and Mäder, P. (2018) Machine learning for image based species identification. *Methods Ecol. Evol.* 9, 2216–2225

19. Creer, S. *et al.* (2016) The ecologist's field guide to sequence-based identification of biodiversity. *Methods Ecol. Evol.* 7, 1008–1018

20. Bush, A. *et al.* (2017) Connecting Earth observation to high-throughput biodiversity data. *Nat. Ecol. Evol.* 1, 0176

21. Balint, M. *et al.* (2018) Environmental DNA time series in ecology. *Trends Ecol. Evol.* 33, 945–957

22. Ruppert, K.M. *et al.* (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* 17, e00547

23. Lahoz-Monfort, J.J. and Magrath, M.J.L. (2021) A Comprehensive Overview of Technologies for Species and Habitat Monitoring and Conservation. *BioScience* 71, 1038–1062

24. Pawlowski, J. *et al.* (2020) Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Mol Ecol* 29, 4258–4264

25. Guisan, A. *et al.* (2017) *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press

26. Elith, J. and Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697

27. Bayraktarov, E. *et al.* (2019) Do Big Unstructured Biodiversity Data Mean More Knowledge? *Front. Ecol. Evol.* 6, 239

28. Leibold, M.A. *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* 7, 601–613

29. Bruelheide, H. *et al.* (2019) sPlot–A new tool for global vegetation analyses. *J. Veg. Sci.* 30, 161–186

30. Abrego, N. *et al.* (2021) Accounting for species interactions is necessary for predicting how arctic arthropod communities respond to climate change. *Ecography* 44, 885–896

31. Carraro, L. *et al.* (2020) Environmental DNA allows upscaling spatial patterns of biodiversity in freshwater ecosystems. *Nat. Commun.* 11, 3585

32. Isaac, N.J.B. *et al.* (2020) Data Integration for Large-Scale Models of Species Distributions. *Trends Ecol. Evol.* 35, 56–67

33. Shelton, A.O. *et al.* (2023) Toward quantitative metabarcoding. *Ecology* 104, e3906

34. Turon, X. *et al.* (2020) From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecol. Appl.* 30, e02036

35. Thomsen, P.F. and Sigsgaard, E.E. (2019) Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecol. Evol.* 9, 1665–1679

36. Jetz, W. *et al.* (2019) Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* 3, 539–551

37. Zurell, D. *et al.* (2018) Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? *Ecography* 41, 1812–1819

38. Blanchet, F.G. *et al.* (2020) Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* 23, 1050–1063

39. Poggiato, G. *et al.* (2021) On the Interpretations of Joint Modeling in Community Ecology. *Trends Ecol. Evol.* 36, 391–401

40. Vellend, M. (2010) Conceptual Synthesis in Community Ecology. *Q. Rev. Biol.* 85, 183–206

41. Leibold, M.A. *et al.* (2022) The internal structure of metacommunities. *Oikos* 2022, e08618

42. Ohlmann, M. *et al.* (2018) Mapping the imprint of biotic interactions on β-diversity. *Ecol. Lett.* 21, 1660–1669

43. Guzman, L.M. *et al.* (2022) Accounting for temporal change in multiple biodiversity patterns improves the inference of metacommunity processes. *Ecology* 103, e3683

44. Banerjee, P. *et al.* (2022) Plant–animal interactions in the era of environmental DNA - A review. *Environ. DNA* 4, 987–999

45. Casey, J.M. *et al.* (2019) Reconstructing hyperdiverse food webs: Gut content metabarcoding as a tool to disentangle trophic interactions on coral reefs. *Methods Ecol. Evol.* 10, 1157–1170

499 46. Kocher, A. *et al.* (2023) Biodiversity and vector-borne diseases: Host dilution
500 and vector amplification occur simultaneously for Amazonian leishmaniases.
501 *Mol. Ecol.* 32, 1817–1831

502 47. Bell, K.L. *et al.* (2017) Applying Pollen DNA Metabarcoding to the Study of
503 Plant–Pollinator Interactions. *Appl. Plant Sci.* 5, 1600124

504 48. Calderón-Sanou, I. *et al.* (2021) Cascading effects of moth outbreaks on
505 subarctic soil food webs. *Sci. Rep.* 11, 15054

506 49. Fišer, C. *et al.* (2018) Cryptic species as a window into the paradigm shift of the
507 species concept. *Mol. Ecol.* 27, 613–635

508 50. Leigh, D.M. *et al.* (2021) Opportunities and challenges of macrogenetic studies.
509 *Nat. Rev. Genet.* 22, 791–807

510 51. Theodoridis, S. *et al.* (2021) Exposure of mammal genetic diversity to mid-21st
511 century global change. *Ecography* 44, 817–831

512 52. Coates, D.J. *et al.* (2018) Genetic Diversity and Conservation Units: Dealing
513 With the Species-Population Continuum in the Age of Genomics. *Front. Ecol.*
514 *Evol.* 6, 165

515 53. Moran, E.V. *et al.* (2016) Intraspecific trait variation across scales: implications
516 for understanding global change responses. *Glob. Change Biol.* 22, 137–150

517 54. Larson, E.R. *et al.* (2020) From eDNA to citizen science: emerging tools for the
518 early detection of invasive species. *Front. Ecol. Environ.* 18, 194–202

519 55. Johnson, M.D. *et al.* (2021) Airborne eDNA Reflects Human Activity and
520 Seasonal Changes on a Landscape Scale. *Front. Environ. Sci.* 8, 563431

521 56. Orlando, L. *et al.* (2021) Ancient DNA analysis. *Nat. Rev. Methods Primer* 1, 14

522 57. Purves, D. *et al.* (2013) Time to model all life on Earth. *Nature* 493, 295–297

523 58. Hagen, O. *et al.* (2021) gen3sis: A general engine for eco-evolutionary
524 simulations of the processes that shape Earth's biodiversity. *PLOS Biol.* 19,
525 e3001340

526 59. Lichstein, J.W. (2007) Multiple regression on distance matrices: a multivariate
527 spatial analysis tool. *Plant Ecol.* 188, 117–131

528 60. Warton, D.I. *et al.* (2015) So Many Variables: Joint Modeling in Community
529 Ecology. *Trends Ecol. Evol.* 30, 766–779

61. Ovaskainen, O. *et al.* (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* 20, 561–576

62. Pollock, L.J. *et al.* (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* 5, 397–406

63. Barraquand, F. *et al.* (2021) Inferring species interactions using Granger causality and convergent cross mapping. *Theor. Ecol.* 14, 87–105

64. Ovaskainen, O. *et al.* (2017) How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proc. R. Soc. B Biol. Sci.* 284, 20170768

65. Delmas, E. *et al.* (2019) Analysing ecological networks of species interactions: Analyzing ecological networks. *Biol. Rev.* 94, 16–36

66. Pilosof, S. *et al.* (2017) The multilayer nature of ecological networks. *Nat. Ecol. Evol.* 1, 0101

67. Tylianakis, J.M. and Morris, R.J. (2017) Ecological Networks Across Environmental Gradients. *Annu. Rev. Ecol. Evol. Syst.* 48, 25–48

68. Gaüzère, P. *et al.* (2022) The diversity of biotic interactions complements functional and phylogenetic facets of biodiversity. *Curr. Biol.* 32, 2093-2100.e3

69. König, C. *et al.* (2021) Scale dependency of joint species distribution models challenges interpretation of biotic interactions. *J. Biogeogr.* 48, 1541–1551

70. Thurman, L.L. *et al.* (2019) Testing the link between species interactions and species co-occurrence in a trophic network. *Ecography* 42, 1658–1670

71. Lahoz-Monfort, J.J. *et al.* (2016) Statistical approaches to account for false-positive errors in environmental DNA samples. *Mol. Ecol. Resour.* 16, 673–685

72. Guillera-Arroita, G. *et al.* (2017) Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods Ecol. Evol.* 8, 1081–1091

73. Tobler, M.W. *et al.* (2019) Joint species distribution models with species correlations and imperfect detection. *Ecology* 100

74. Devarajan, K. *et al.* (2020) Multi-species occupancy models: review, roadmap, and recommendations. *Ecography* 43, 1612–1624

75. Diana, A. *et al.* (2022) eDNAPlus: A unifying modelling framework for DNA-based biodiversity monitoring. arXiv:2211.12213, https://doi.org/10.48550/arXiv.2211.12213

76. Rhinehart, T.A. *et al.* (2022) A continuous-score occupancy model that incorporates uncertain machine learning output from autonomous biodiversity surveys. *Methods Ecol. Evol.* 13, 1778–1789

77. Dietz, T. *et al.* (2003) The Struggle to Govern the Commons. *Science* 302, 1907–1912

78. Ji, Y. *et al.* (2022) Measuring protected-area effectiveness using vertebrate distributions from leech iDNA. *Nat. Commun.* 13, 1555

79. Laschever, E. *et al.* (2023) Next Generation of Environmental Monitoring: Environmental DNA in Agency Practice. *Columbia J. Environ. Law* 48, 51

80. Arribas, P. *et al.* (2021) Connecting high-throughput biodiversity inventories: Opportunities for a site-based genomic framework for global integration and synthesis. *Mol. Ecol.* 30, 1120–1135

81. Mair, L. *et al.* (2021) A metric for spatially explicit contributions to science-based species targets. *Nat. Ecol. Evol.* 5, 836–844

82. Bohmann, K. and Lynggaard, C. (2023) Transforming terrestrial biodiversity surveys using airborne eDNA. *Trends Ecol. Evol.* 38, 119–121

83. Clare, E.L. *et al.* (2022) Measuring biodiversity from DNA in the air. *Curr. Biol.* 32, 693–700

84. Bohmann, K. *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29, 358–367

85. Taberlet, P. *et al.* (2018) *Environmental DNA*, 1, Oxford University Press

86. Ratnasingham, S. and Hebert, P.D.N. (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE* 8, e66213

87. Ficetola, G.F. and Taberlet, P. (2023) Towards exhaustive community ecology via DNA metabarcoding. *Mol. Ecol.* DOI: 10.1111/mec.16881

88. Mathon, L. *et al.* (2021) Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Mol. Ecol. Resour.* 21, 2565–2579

89. Kelly, R.P. *et al.* (2019) Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Sci. Rep.* 9, 12133

90. Luo, M. *et al.* (2023) Extracting abundance information from DNA-based data. *Mol. Ecol. Resour.* 23, 174–189

91. Griffin, J.E. *et al.* (2020) Modelling Environmental DNA Data; Bayesian Variable Selection Accounting for False Positive and False Negative Errors. *J. R. Stat. Soc. Ser. C Appl. Stat.* 69, 377–392

92. Williamson, B.D. *et al.* (2022) A multiview model for relative and absolute microbial abundances. *Biometrics* 78, 1181–1194

93. Somervuo, P. *et al.* (2017) Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. *Methods Ecol. Evol.* 8, 398–407

94. Zito, A. *et al.* (2023) Inferring taxonomic placement from DNA barcoding aiding in discovery of new taxa. *Methods Ecol. Evol.* 14, 529–542

95. Pichler, M. and Hartig, F. (2021) A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods Ecol. Evol.* 12, 2159–2173

96. Wilkinson, D.P. *et al.* (2021) Defining and evaluating predictions of joint species distribution models. *Methods Ecol. Evol.* 12, 394–404

97. Biggs, J. *et al.* (2015) Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (Triturus cristatus). *Biol. Conserv.* 183, 19–28

98. Trujillo-González, A. *et al.* (2021) Considerations for future environmental DNA accreditation and proficiency testing schemes. *Environ. DNA* 3, 1049–1058

99. Natural England (2019) *A Framework For District Licensing Of Development Affecting Great Crested Newts.* Natural England Technical Information Note TIN176, https://publications.naturalengland.org.uk/publication/5106496688095232

100. Bush, A. *et al.* (2023) Systematic Nature Positive Markets. bioRxiv, 2023.02.13.528257, https://doi.org/10.1101/2023.02.13.528257

Glossary

Biotic interaction: a direct (e.g. competitive, mutualistic, trophic) interaction between individuals of two different taxa

626 **Biodiversity monitoring**: the process of generating information about the spatio-
627 temporal distribution of biodiversity. The produced data is often represented as a
628 community matrix (see below).

629 **Community inventory** (also: Biodiversity inventory): a list of species occurring in a
630 particular place and time. Conventional inventories often target a particular species
631 group.

632 **Community matrix:** a matrix consisting of many community inventories, traditionally
633 with rows = inventory number (sites or time), and columns = species or taxa,
634 characterizing presence, presence-absence, abundance or biomass for each
635 species / site combination.

636 **Cryptic species:** species that are morphologically indistinguishable but genetically
637 distinct and reproductively isolated and can thus only reliably be identified with
638 molecular analyses.

639 **Environmental DNA (eDNA):** DNA isolated from environmental samples, including
640 both extraorganismal (trace) and organismal eDNA. For example, bulk-arthropod
641 samples contain both organismal eDNA from arthropods and trace eDNA from
642 vertebrates (e.g. blood, feces, skin).

643 **Exact-sequence variants (ESVs):** unique DNA sequences that are identified from
644 high-throughput sequencing. Unlike more traditional operational taxonomic units
645 (OTUs, see below), which cluster non-identical but similar sequences, ESVs
646 describe identical nucleotide sequences.

DNA barcoding: species identification using a short section of DNA from a specific gene or genes, which is mapped against a barcoding reference database

joint Species Distribution Model (jSDM): a statistical model that describes a vector of community (multi-species) presences or abundances as a function of abiotic, biotic or spatial predictors (like an SDM) and an additional component, which consists of residual covariances between the modeled species, describing positive or negative species associations.

Metadata: in general, data describing other data. In the context of this paper, we include in this definition all data that complement the primary community observations.

Novel community data: large community datasets generated by automated pipelines such as eDNA sequencing and electronic sensors (e.g. bioacoustics sensors or visual sensors such as camera traps).

Operational Taxonomic Unit (I): a group of haplotypes that are clustered together based on their sequence similarity to form distinct taxonomic entities, typically species.

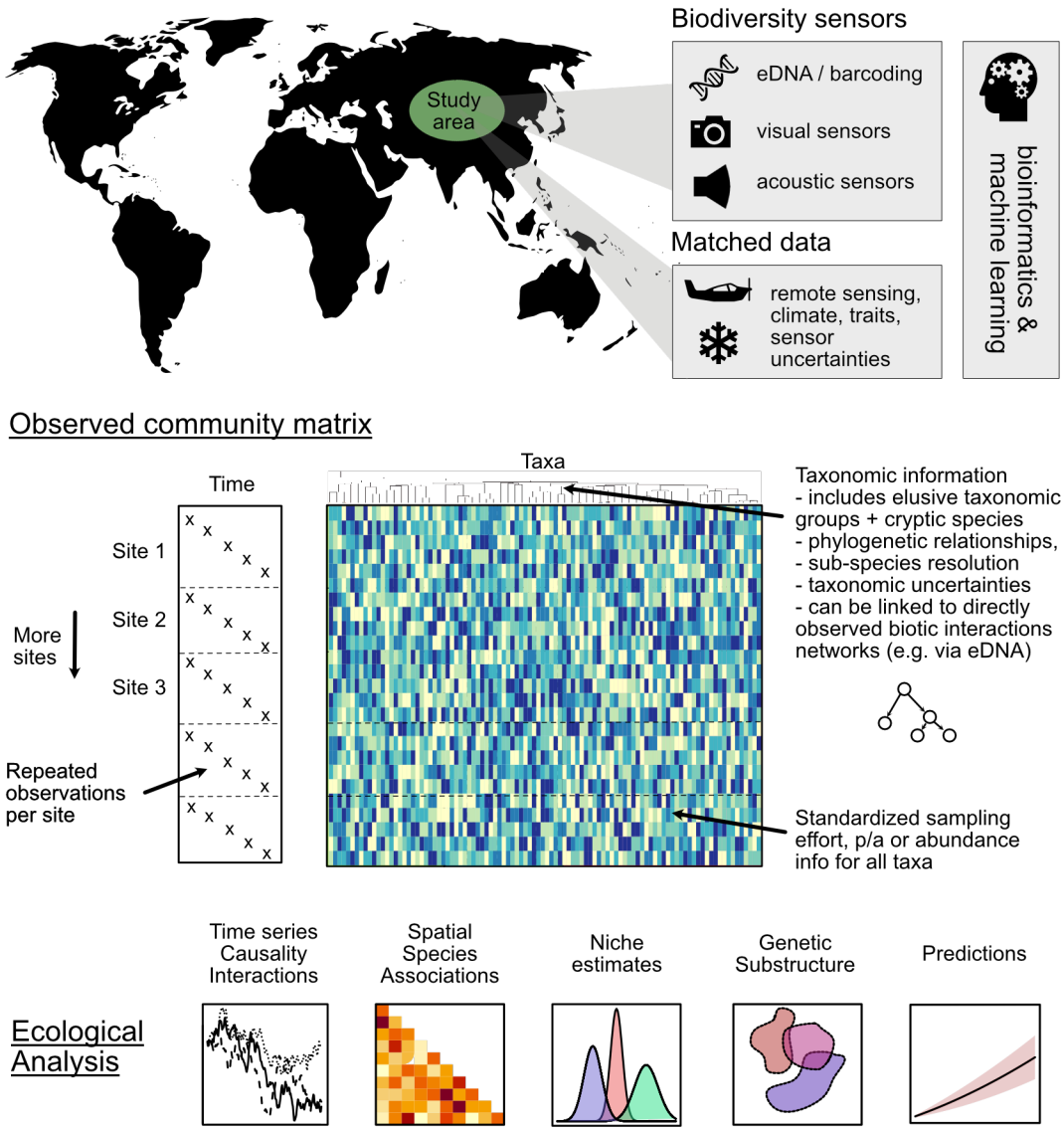
Passive acoustic monitoring: deployment of acoustic sensors in the field to detect sounds created by wildlife and the surrounding (soundscape). This data can be processed by experts or machine learning methods to classify the sounds of specific species or communities.

Pipeline: a series of computational and analytical steps to process and analyze raw sensor data such as sequencing data, acoustic observations, or pictures.

Species Distribution Model (SDM): a statistical model that relates species presence or abundance data to a set of abiotic, biotic or spatial predictors.

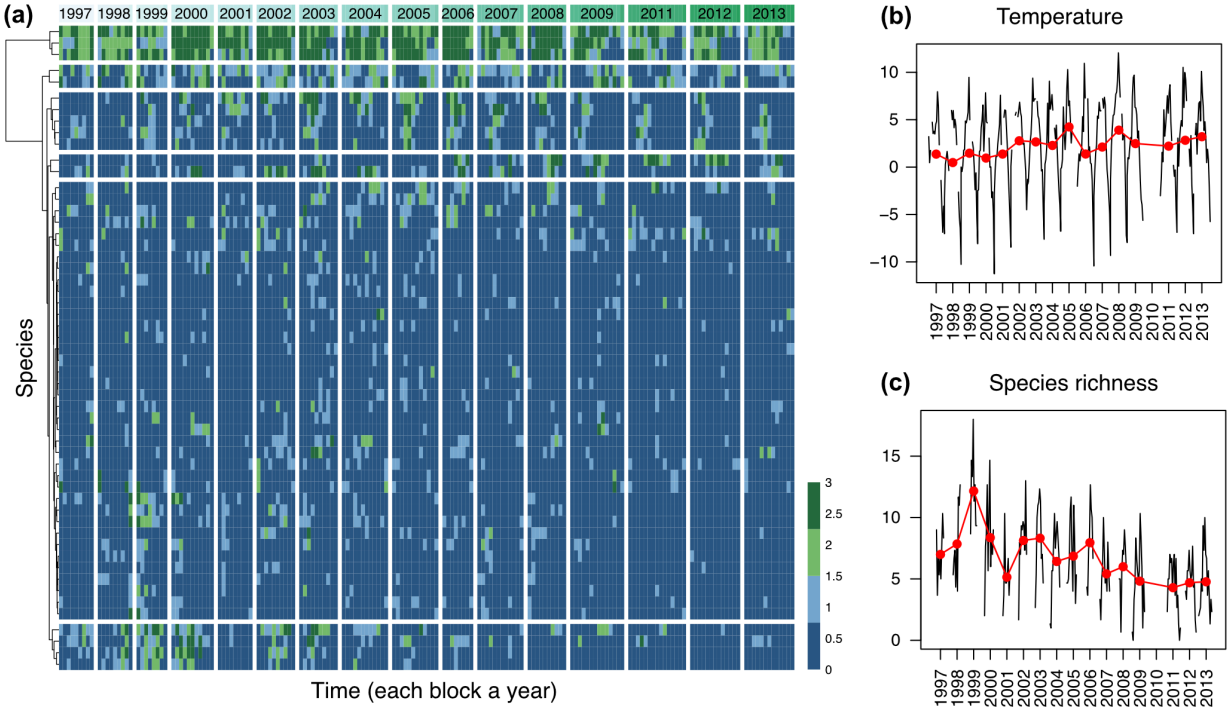
Species association: a correlation or association of occurrence, abundance, or distribution of two taxa, which can be due to biotic interactions, (missing) environmental covariates, distributional disequilibrium, and other reasons.

Visual sensors: we use visual sensors as an umbrella term for all optical sensors that can be used for species identification. This includes photos, e.g. from camera traps, videos and potentially also visual information from remote sensing, in particular from drones.



678

679 **Fig. 1:** Novel biodiversity sensors generate detailed community inventories as well
680 as rich metadata. If replicated in space and time, this gives rise to novel community
681 data. This novel community, represented in the center of the figure, is more
682 information-dense in many dimensions beyond spatial replicates, including time,
683 taxonomic relationships, and interaction information. As a result, these data allow for
684 a richer set of ecological analyses than conventional community inventories.



685

686 **Fig. 2:** Abrego et al. [30] analyzed a 16-year, weekly community time series of
687 arthropod community dynamics from Greenland, resolved to the species level by
688 eDNA mitogenome mapping. Panel a) shows the species x time community matrix,
689 with cell colors indicating the number of traps out of 3 in which the species was
690 detected at each point in time. During the study period, temperature increased by
691 2°C and arthropod species richness halved (panels b respectively c, reprinted from
692 [30]). In their analysis of the data, Abrego et al. show that abiotic variables alone are
693 insufficient to predict species responses, but with species interactions included, the
694 predictive power of the model improves. Trophic cascades thereby emerge as
695 important in structuring biodiversity response to climate change. The study
696 emphasizes the potential of eDNA data to generate high-resolution community time
697 series and thus to understand the complex interplay of biotic and abiotic effects in

698 climate change impacts. The analytical tools used to reach these conclusions are
699 explained in Box 2.

700 Box 1: An overview of the eDNA pipeline

701 All species shed DNA into the environment. We refer to this DNA isolated from
702 environmental substrates, even air [82,83], as eDNA [24,84,85]. eDNA can either be
703 sequenced *en masse* and processed *in-silico* to find taxonomically informative
704 sequences (metagenomics) or read after targeted amplification of taxonomically
705 informative sequences in the laboratory (metabarcoding). The resulting DNA
706 sequences ('reads') are typically first clustered to **operational taxonomic units**
707 (OTUs) and then compared to DNA-barcode reference databases to assign
708 taxonomies [87].

709 Although the eDNA pipeline can in principle detect all cellular organisms, the
710 achieved taxonomic coverage in current eDNA studies is limited by the physical
711 collection of eDNA material, by the molecular methods used, and, for taxonomic
712 assignment beyond OTUs or ESVs, by the availability of suitable reference
713 databases [87]. Future methods will likely expand taxonomic coverage, but even
714 existing methods enable the standardized detection of many species across trophic
715 groups, including cryptic, difficult-to-observe, small, and low abundance species,
716 from easily collected samples.

717 Practical challenges in using eDNA include the high diversity of different
718 bioinformatic pipelines for curating, cleaning and clustering eDNA sequences (but

719 see [88]), as well as dealing with eDNA-specific sampling and detection errors
720 (Table I, see also [15,75,89,90]). For example, stochasticity and sample-equalization
721 steps in laboratory pipelines can obscure the expected positive relationship between
722 the eDNA biomass and the resulting number of reads, but adding a DNA spike-in to
723 each sample can help to recover this relationship [75,90]. Moreover, sample
724 contamination can result in false-positive errors. Good practice limits such events to
725 be rare and weak, letting false positives be identified [91].

726 A further challenge with eDNA data is that the number of eDNA ‘reads’ per individual
727 depends in part on unknown, species-specific rates of release, degradation, and
728 PCR efficiency ('species effects', Table I, see also [16]). As a result, eDNA reads are
729 in general not proportional to species abundances or biomass. However, if (1) eDNA
730 release, degradation, and PCR efficiency are approximately constant across
731 samples, and (2) pipeline stochasticity is accounted for (via spike-in estimated
732 offsets), then across-sample change in reads for each species are proportional to
733 across-sample changes in that species’ abundance [33,75,90,92].

734 Finally, taxonomic assignment can have errors or uncertainties due to incomplete
735 reference databases and variation across species in genetic diversity. Ideally, such
736 errors are accounted for by dedicated statistical methods. For example, Bayesian
737 algorithms can be trained to estimate the degree of sequence similarity required to
738 assign membership to a given rank within a given taxon [93,94].

739 **Table I:** The two stages of DNA-based surveys and the sources of false-negative
740 error, false-positive error, and row, column, and cell effects in the output sample x
741 species table (adapted from [75]).

Stage 1 - eDNA biomass collection		Analogues in conventional surveys
Species effects	Every sample collects a certain amount of eDNA biomass of each species, which is proportional to the species' biomass available at the site. However, the proportionality constant is marker- and species-specific and is unknown, since rates of DNA release, 'catchability', and degradation differ across species and physiological states (a 'column' effect).	Species differ in their detectability by human observers or by trapping bias.
Noise	The amount of eDNA biomass collected per species varies stochastically among samples collected at the same site and time (a 'row' effect), including outright collection failure (false negatives).	Imperfect detection of species, false negatives
Error	It is possible for traces of eDNA from elsewhere to contaminate a sample (false positives).	No analogue in conventional surveys
Stage 2 - eDNA lab + bioinformatics pipeline		Analogues in conventional surveys
Species effects	Species differ in extraction efficiency, gene copy number, and PCR amplification efficiency, causing the relationship between input eDNA amount and number of output sequence reads to be species-specific (a 'column' effect).	Species differ in their detectability by human observers or by trapping probabilities.
Pipeline effect	PCR stochasticity, normalization steps, and the passing of small aliquots of liquid along the lab pipeline add stochasticity to the total number of output reads per sample replicate (a 'row' effect), including outright detection failure (false negatives).	No analogue in conventional surveys
Noise	On top of species and pipeline effects, there is additional noise in the number of reads per species, sample and/or technical replicate (a 'cell' effect).	No analogue in conventional surveys
Contamination Error	It is possible for traces of eDNA from one sample to contaminate other samples (false positives).	No analogue in conventional surveys
Barcoding errors	Incorrect delimitation of sequence variation leading to incorrect taxonomic lumping or splitting; or incorrect species identification because the sequence is wrongly assignment to a taxonomy (paired false-negative / false-positive errors)	Incorrect lumping of cryptic species or incorrect splitting of a single species; or species misidentification resulting in paired false-

		negative / false-positive errors

742

Box 2: jSDMs as a tool to model novel community data

In recent years, joint species distribution models (jSDMs) have emerged as the main extension of classical species distribution models for the analysis of community data [60–62]. The key difference between SDMs and jSDMs is that while the former can also model communities, they do so by describing each species individually (stacked SDMs).

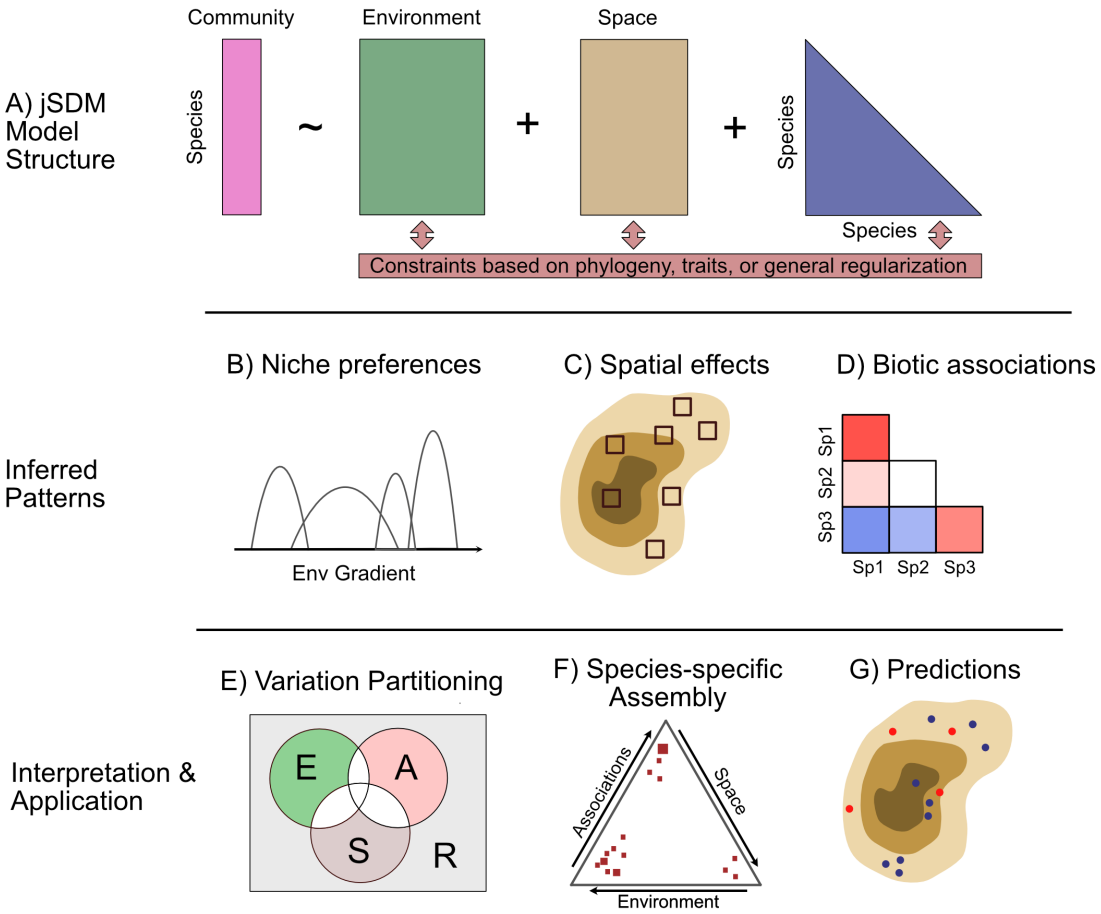
A jSDM, however, is a true community model because, additional to the environmental responses of each species, it includes a species-species covariance component. This covariance models species associations, meaning the tendency of species pairs to co-occur more or less frequently than one would expect based on their species-specific environmental preferences alone (see Fig. I).

The basic jSDM structure can be extended to include additional correlations in species' niche estimates via phylogeny or traits, and spatial predictors. jSDMs can also be extended to fit spatio-temporal data, which allows one to consider additionally asymmetric associations [63,64]. Due to their complex likelihood, jSDMs are often challenging to fit, and several numeric strategies, including the latent-variable approximation (e.g. [60]) and Monte-Carlo approximations [95], have been proposed to make these models scalable to large community data.

The interpretation of the species associations inferred by jSDMs has been the subject of considerable debate in the field. We view it now as accepted that species *associations* are not necessarily caused by biotic *interactions* (e.g. [38]; but see

764 [37]). Among other things, this implies that a jSDM will typically not improve the
765 estimation of the fundamental niche [39]. Nevertheless, the ability to partition the
766 community signal into the three classical components of environment, space, and
767 association (Fig. B2), which can further be broken down to sites (communities) and
768 species (i.e. the 'internal structure', see [41] and Fig. B2), provides a rich framework
769 for analyzing spatial community data. Moreover, if some species can be easily
770 observed, conditioning on their presence using jSDMs can also improve predictions
771 [96], which may be relevant for management.

772



773

Fig. I: An overview of structure, inferred patterns and interpretation of a jSDM. A) A possible jSDM structure, predicting community composition based on environment, space and species-species covariance. B) Environmental effects show niche preferences C) Spatial effects show spatial clustering of species D) Species-species covariance shows species associations. E) An ANOVA of the entire jSDM (A) can partition community variation into Environment, Space, Associations and Residual components. F) This can further be broken down by species or sites [41], so that we can see the relative importance of the three components to individual species and sites. G) If particular presences are known (red), we can condition on them to improve predictions [96].

Box 3: An eDNA-enabled biodiversity offset market

One example of institutional redesign enabled by eDNA is the District Licensing Market for the great crested newt (*Triturus cristatus*), a protected species in the UK. Developers are required to survey for the newt when their plans may affect ponds, and to respond to newt detections by paying for mitigation measures. Traditional surveys require at least four visits per pond during the short breeding season, using multiple methods that are only effective at night. Following a study [97] showing that a single eDNA water survey could detect the newt with the same sensitivity as traditional surveys (i.e. eDNA detections are *high-quality* and *granular*), the government authorized newt eDNA surveys in 2014, and a private market for eDNA surveys, audited with proficiency tests, grew to provide the *infrastructure* for *timely* and *trustworthy* information [98].

796 The switch to eDNA surveys increased survey efficiency, but the UK's reactive
797 (mitigate-after-impact) approach was initially left in place. Mitigation measures, such
798 as translocation, can take over a year, with associated costs. In 2018, the UK
799 government took further advantage of eDNA's efficiency by implementing an
800 *institutional redesign* with the District Licensing scheme, in which the ponds across
801 one or more local planning authorities are systematically surveyed with eDNA [99].
802 The data is then used to fit a species distribution model, which is made into an
803 *understandable* map of discrete background risk zones for the newt (Fig. I). Builders
804 can meet their legal obligations at any time by paying for a license, the cost of which
805 depends on the size of their site, the background risk zone, and the number of
806 ponds affected.

807 The fees from these licenses are mainly used towards the proactive creation and
808 long-term management of compensation habitat including ponds with a one-to-four
809 impact-to-gain ratio. The compensation habitat is directed toward Strategic
810 Opportunity Areas that account for planning-authority building aspirations (*political*
811 *bargaining*). *Enforcement* is through the same processes that apply to all planning
812 permissions. Both the UK government and a private-public-NGO partnership run
813 versions of District Licensing markets, which together have reported creating
814 hundreds of new ponds and associated habitat. In the future, it might be possible to
815 effect a further institutional redesign by exploiting the multi-species information in the
816 pond water samples to move to multi-species conservation planning and offset
817 markets [100].

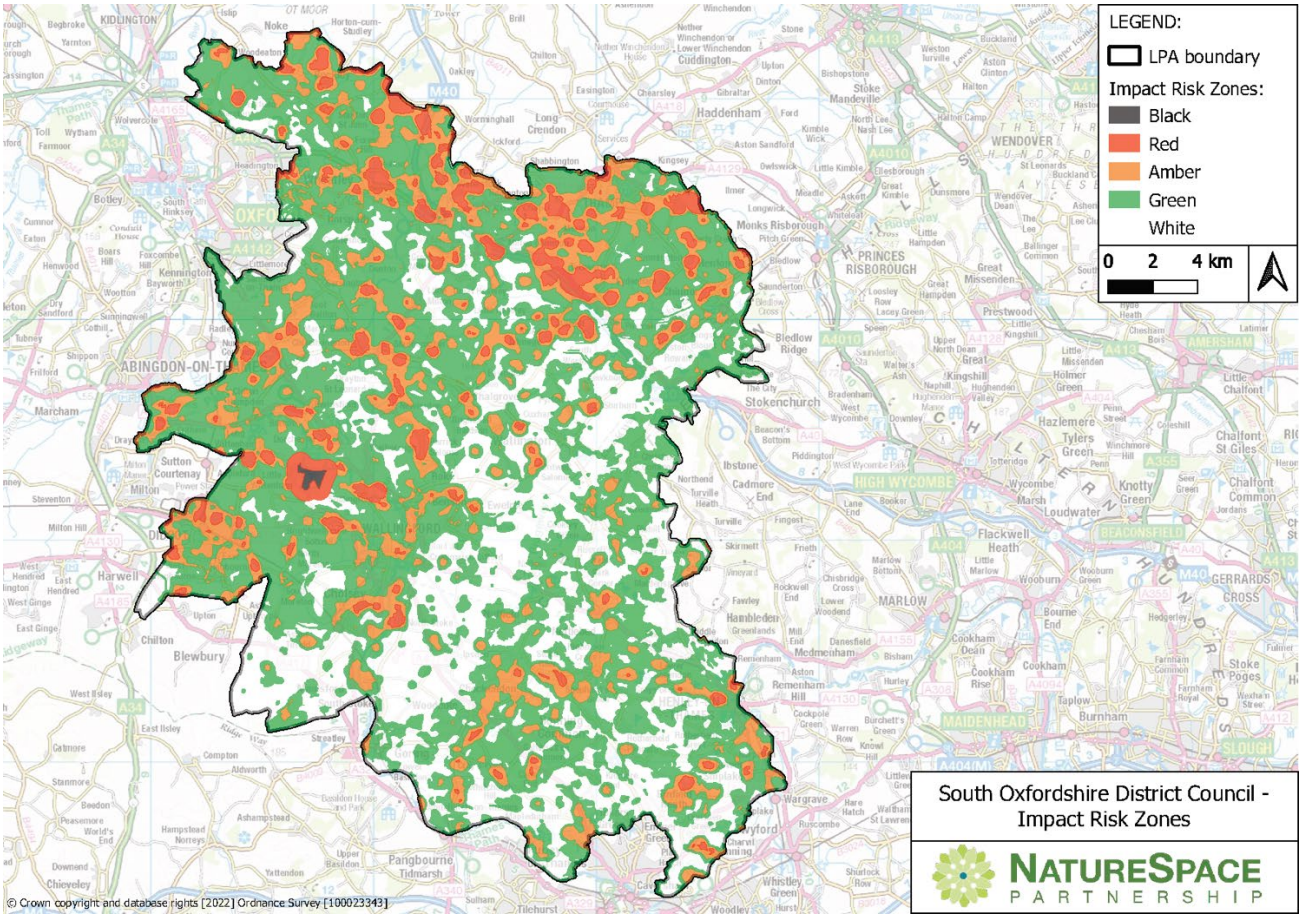


Fig. I: Risk zone map for great crested newt (*Triturus cristatus*) in one Local Planning Authority (LPA). Reprinted with permission from NatureSpace Partnership.