From Scalable Biodiversity Measurement to Credible Biodiversity Metrics

Authors: Douglas W. Yu[1,2]*, Fabian Ewald Fassnacht[3], Tone Birkemoe[4], Corrado Di Maria[5], Emiliya Lazarova[5], Francesca Cuomo[6], Giovanna Michelon[7], Hans Ole Ørka[4], Terje Gobakken[4], Viorel D. Popescu[8,9], Maximilian Pichler[10]

Affiliations:

[1]School of Biological Sciences, University of East Anglia, Norwich, UK
[2]Yunnan Key Laboratory of Biodiversity and Ecological Security of Gaoligong Mountain, Kunming Institute of Zoology, Kunming, China
[3]Institute of Geomatics, BOKU University, Vienna, Austria
[4]Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway
[5]School of Economics, University of East Anglia, Norwich, UK
[6]Norwich Business School, University of East Anglia, Norwich, UK
[7]Department of Economics and Management, University of Padova, Padova, Italy
[8]School of Professional Studies & Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, USA
[9]Center for Environmental Research (CCMESI), University of Bucharest, Bucharest, Romania
[10]Theoretical Ecology, University of Regensburg, Regensburg, Germany
*Corresponding author. Email: douglas.yu@uea.ac.uk

Nullius in verba – Take no one's word for it (Motto of the Royal Society)

Abstract

Governments struggle to develop effective policies to counter the decline of species and ecosystems. An obstacle to command-and-control and incentive-based mechanisms is that biodiversity is costly to measure, creating an information asymmetry in which firms and governments are incentivised to withhold information on adverse impacts. Using a principal-agent model, we show that credible reporting requires biodiversity measurement to satisfy four criteria: *low marginal cost, low dispersion, sufficient information content,* and *parsimony*.

Though not yet deployable, a route towards meeting these criteria is emerging: the integration of deep-learning species distribution models (DL-SDMs) with remote sensing, proximal sensing, novel community data, and citizen science, which we term *Scalable Biodiversity Measurement*.

We assess DL-SDM readiness against the Mitigation Hierarchy (Avoidance, Minimisation, Remediation, Offsetting), which spans the full range of actions disclosed in sustainability reporting. The components for auditing avoidance, spatial minimisation, and conservation offsets now exist, pending investment in infrastructure and training data. Operational minimisation and remediation remain immature, requiring advances in causal attribution.

We propose a roadmap to scale up credible biodiversity metrics: (1) investment in large, standardised training datasets; (2) a transparent political process to compress high-dimensional outputs into parsimonious metrics; and (3) deeper integration of biodiversity science with mechanism-design economics.

Keywords

asymmetric information, biodiversity credits, biodiversity net gain, Corporate Sustainability Reporting Directive CSRD, ecosystem assessment, ecosystem condition, ecosystem health, greenwashing, information asymmetry, nature positive, nature-related financial disclosures, no net loss, payments for ecosystem services, principal-agent model, sustainability reporting, systematic conservation planning, Taskforce on Nature-related Financial Disclosures TNFD

Key points

1. Biodiversity is costly to measure, creating an information asymmetry in which firms and governments are incentivised to hide adverse impacts.
2. Credible reporting requires biodiversity measurement to satisfy four criteria: *low marginal cost, low dispersion, sufficient information content,* and *parsimony*.
3. A route to meeting these criteria is emerging: integrating deep-learning species distribution models (DL-SDMs) with remote sensing, proximal sensing, novel community data, and citizen science.
4. The components for auditing avoidance, spatial minimisation, and conservation offsets now exist, whereas operational minimisation and remediation remain immature.
5. We propose a roadmap: large standardised training datasets, a transparent political process to compress outputs into parsimonious metrics, and deeper integration with mechanism-design economics.

Information asymmetry and the route to credible biodiversity metrics

The Kunming-Montreal Global Biodiversity Framework (KMGBF) has formally adopted a broad suite of mechanisms, including national accounting (Target 14), corporate impact disclosures (Target 15), consumer eco-labelling (Target 15), green subsidies (Target 18), and biodiversity offsets and credits (Target 19), to incentivise conservation and restoration and close the global biodiversity financing gap(Convention on Biological Diversity, 2022; Deutz et al., 2020). However, all of these mechanisms depend on the truthful disclosure of biodiversity impacts by firms and governments to be effective(Carter et al., 2026; Deutz et al., 2020; Ducros & Steele, 2022; Elliot et al., 2024; Sutherland et al., 2026; Wunder et al., 2025). But biodiversity is “notoriously difficult” to measure at scale(Elliot et al., 2024; Marshall et al., 2020, 2024; Nicholson Thomas et al., 2025). This observability problem incentivises firms and governments to conceal adverse impacts, creating an information asymmetry that weakens the credibility of biodiversity disclosures and the policies built upon them(Carter et al., 2026; Swinfield et al., 2024; zu Ermgassen et al., 2026) (Figure 1).

Information asymmetry means that one side in a transaction knows more than the other; in Akerlof’s(1970) classic example, sellers know whether their used cars are faulty, but buyers do not, leading buyers to offer lower prices or walk away. Likewise, even when firms are legally obligated to disclose their biodiversity impacts, the impacts are spread across operations, supply chains, and land-use footprints that consumers, investors, and regulators cannot easily observe from disclosed information alone. Without effective third-party verification, firms can claim nature-positive performance without incurring the cost of achieving it, and markets and regulators cannot reliably distinguish genuine conservation from greenwashing. Policies and financial instruments thus risk rewarding the appearance of good performance rather than the biodiversity outcomes themselves, while governments, investors, and consumers, unable to identify credible claims, walk away from biodiversity investment. This dynamic is not hypothetical; analogous credibility failures in carbon markets contributed to a sharp decline in demand when the quality of carbon credits was called into question(Forest Trends’ Ecosystem Marketplace, 2024; West et al., 2023).

Figure 1. The route to credible biodiversity metrics. The Problem. Auditors have limited budgets, letting adverse impacts go unreported. To be credible, biodiversity metrics must meet four criteria. **The Potential Solution**. For terrestrial ecosystems, deep-learning species distribution models combine remote sensing and scalable species sensing into large training datasets. Optional proximal-sensing adds data at selected sites, raising local accuracy at higher marginal cost. Trained models predict many species distributions, which are reduced to one or a few mappable biodiversity metrics. **Audit Readiness Level**. The technological components for auditing avoidance, spatial minimisation, and conservation offsets now exist, whereas operational minimisation and remediation remain immature. **Roadmap**. Routine scalable biodiversity measurement requires large training datasets, agreement on dimension-reduction algorithms to produce a parsimonious set of metrics, and integration of biodiversity science with mechanism-design economics.

THE PROBLEM

INFORMATION ASYMMETRY

FIRM

Knows its own activities, can choose to measure its own impact, can choose what to report

AUDITOR

Can allocate some of its limited budget to audit (with uncertainty) a firm's impact claims

CREDIBLE METRICS REQUIRE

LOW MARGINAL COST

Credible threat of audit

SUFFICIENT INFORMATION CONTENT

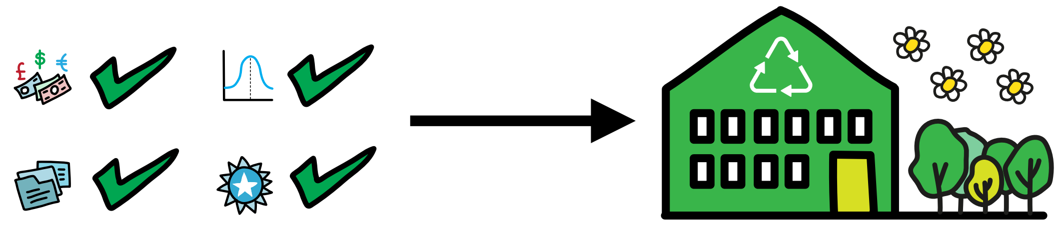
Enough legitimacy to condition consumer and investor behaviour

LOW DISPERSION

Audits carry evidentiary force

PARSIMONY

Prevents cherry-picking

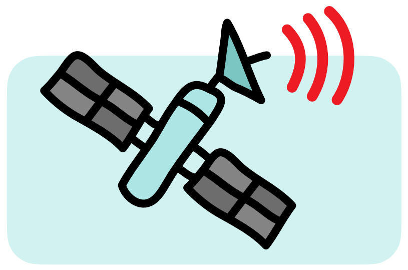


All four required

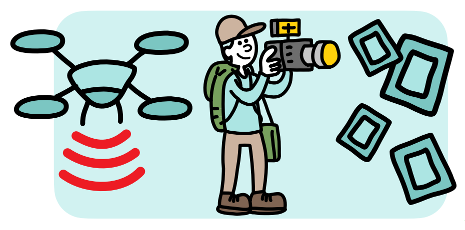
Any gap enables greenwashing

The first three criteria emerge directly from a principal-agent model of the interaction between an auditor and a reporting firm. Parsimony follows narratively by allowing firms to select from a menu of metrics.

THE POTENTIAL SOLUTION - SCALABLE BIODIVERSITY MEASUREMENT



Remote Environmental Sensing



Optional proximal sensing of additional environmental covariates and/or species in selected sites

Scalable Species Sensing

ENV COVARIATES

SITES

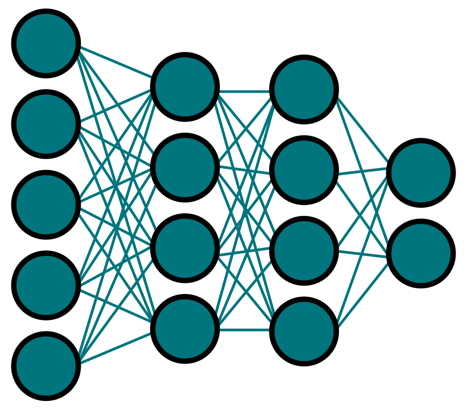
-3	84	62	-12	-4	72
12	92	71	8	11	81
-8	78	58	-9	8	68
22	99	79	15	18	85
5	89	65	-6	3	76
-5	81	69	-18	14	79

SPECIES

SITES

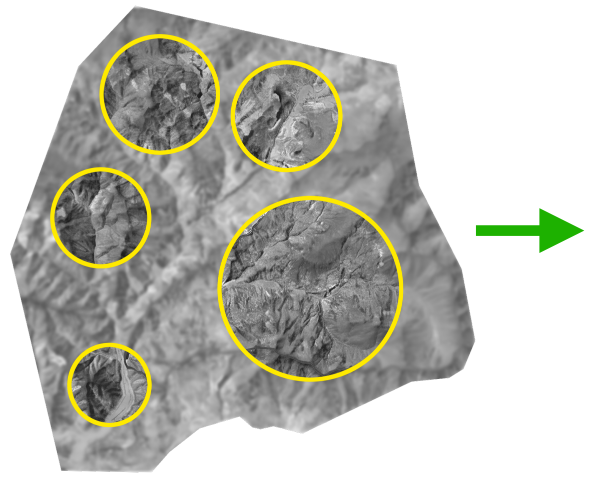
1	0	1	0	1	0
1	0	1	0	1	0
1	0	1	0	1	0

DL-SDM (deep-learning-based species distribution model)



A DL-SDM takes as input continuous environmental measures, including encoded remote-sensing imagery, and species observations from a finite set of locations.

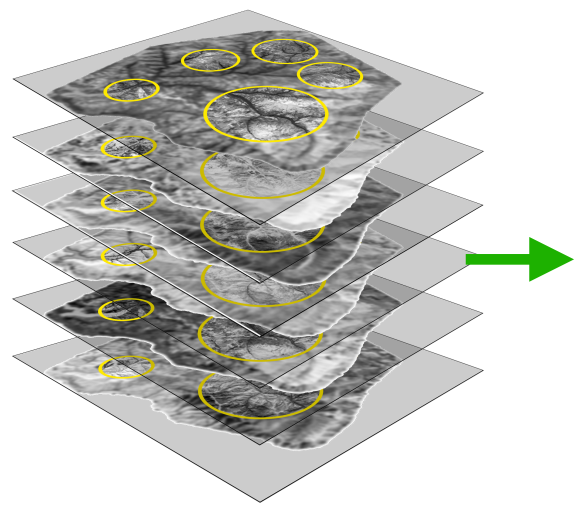
A continuous species distribution. Higher accuracy in sites with proximal sensing data, but at higher marginal cost



The trained DL-SDM model predicts the presence of each species continuously across a landscape.

Optional additional proximal-sensing data at selected locations (yellow circles) can increase predictive accuracy.

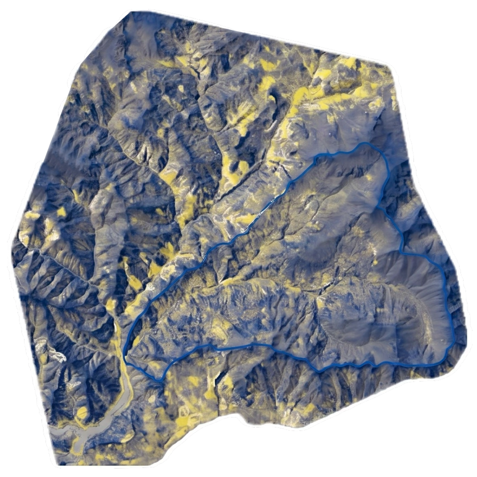
Multiple continuous species distributions



The output of a DL-SDM is a stack of species distribution maps.

Alternative DL-models directly predict and map summary metrics like forest types.

A bottom-up biodiversity metric

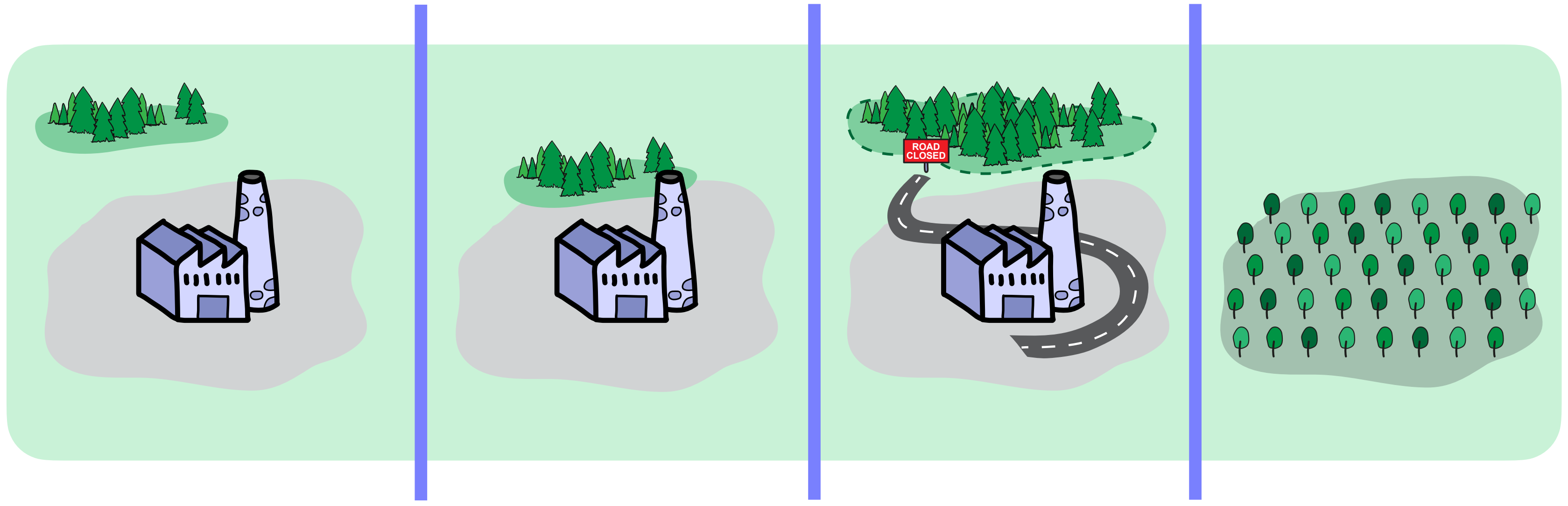


The high-dimensional output of a DL-SDM is reduced to a one-dimensional metric to support decision-making and prevent cherry-picking. Here, the metric is site irreplaceability (yellow=higher SI=areas to avoid).

AUDIT READINESS LEVEL

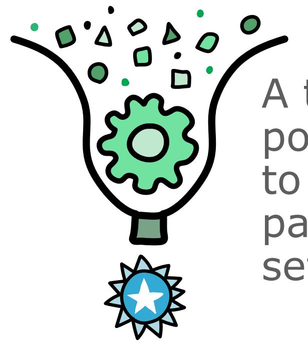
Mitigation Hierarchy actions

Avoidance (High readiness)	Spatial Minimisation (High readiness)	Conservation Offsets (Medium readiness)	Operational Minisation and Remediation (Low readiness)
Avoid impacts by sitinginfrastructure and extraction (e.g. logging) fully away from high-conservation-value sites	Reduce biodiversity impacts by shifting infrastructure toward areas of lower conservation value	Generate biodiversity gains away from the impacted site by conserving sites under threat. Requires counterfactual.	Adjust activities to reduce harm or recover biodiversity at the impacted site. Requires causal attribution and/or counterfactual.

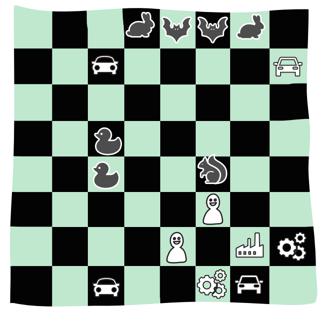


ROADMAP

Capital investment to create large training datasets in every biome-region combination



A transparent political process to agree on a parsimonious set of metrics



Deeper integration of biodiversity science with mechanism-design economics

To make biodiversity disclosure credible, the central challenge is to generate measurements that can support truthful reporting under real-world conditions of limited audit capacity and strategic behaviour. We illustrate this challenge using the EU's new Corporate Sustainability Reporting Directive(CSRD, 2022) as a case study. Annual CSRD disclosures are mandatory for large firms with headquarters or major operations in the EU(de Groen et al., 2022), requiring them to report both their *impacts* on people and the environment and the ways in which environmental or social conditions could create *risks* or *opportunities* for the firm itself. Yet the CSRD also reveals the limits of current biodiversity measurement and assurance systems. As we show in *Supplementary Information 1*, the biodiversity disclosure requirements are highly vulnerable to greenwashing. Firms retain substantial discretion over which biodiversity impacts they consider material and therefore disclosable, which metrics they use to measure those impacts, and which, if any, policies and targets they adopt in response. At the same time, the scale of future reporting – expected to be ~10,000 reports annually – makes truly independent verification difficult to imagine using conventional auditing approaches.

Where effective auditing is absent, firms are incentivised to under-report harms and overstate positive performance(Blanco-Zaitegi et al., 2024; Boiral & Heras-Saizarbitoria, 2020; Swinfield et al., 2024; Walker et al., 2009). The result is low credibility of reporting and a weak foundation for designing regulations and financial incentives that generate genuine corporate demand for conserving nature.

We formalise these findings in a game theory model of the interaction between an auditor (a *principal*) and a reporting firm (an *agent*) (*Supplementary Information 2*). This model demonstrates that truthful reporting is only likely to emerge when biodiversity measurements satisfy four criteria:

1. **Low Marginal Cost**: If biodiversity measurement remains expensive, third-party verification will be too sparse to create a credible threat of audit(Kedward et al., 2023; Kujala et al., 2022; Vardon & Lindenmayer, 2023; Walker et al., 2009). This problem already affects sustainability assurance, which is criticised for failing to demonstrate meaningful, independent verification(Boiral & Heras-Saizarbitoria, 2020).
2. **Low Dispersion**: Precision matters because auditors must be able to distinguish deliberate misreporting from natural variability and measurement noise. If estimates are known to be very noisy, even a formally independent audit may carry little evidentiary force.
3. **Sufficient Information Content**: Disclosures must capture enough information about biodiversity status and change to convince auditors, informed consumers, regulators, and financial institutions that reported impacts are ecologically meaningful and correctly assessed(Carter et al., 2026; Wunder et al., 2025). In practice, this implies information that is timely, granular in space, time and

taxonomy, and explicit about measurement uncertainty(Dietz et al., 2003). For some applications, it also requires causal attribution.

4. **Parsimony**: The larger the menu of permitted metrics, the easier it becomes for firms to cherry-pick metrics that cast performance in the most favourable light while remaining difficult to verify or that are ambiguous over clear-but-negative ones, as seen in drug trials(Dahm et al., 2018). Restricting reporting to one or a very small number of metrics reduces opportunities for cherry-picking, increases understandability, and improves the performance of voluntary biodiversity credit markets(Burgess et al., 2024; Ducros & Steele, 2022; Elliot et al., 2024; Layman et al., 2024; Nature Positive Initiative, 2026; Rossberg et al., 2024; Walker et al., 2009).

The first three criteria emerge directly from comparative statics of the formal model (*Supplementary Information 2*). Parsimony follows narratively by allowing firms to select from a menu of metrics. The four criteria together define what we call **credible biodiversity metrics**, and they provide a practical framework for judging whether emerging biodiversity measurement systems are ready to support high-stakes governance applications.

It is instructive to contrast our model-derived criteria to zu Ermgassen et al.'s(2026) five criteria for "scientifically credible nature markets". Three of their criteria (*ensure proxy correlates with outcome*, *assume nonadditionality unless proved otherwise*, *assess and manage leakage*) fall under our criterion 3 for sufficient information content. The other two (*ensure independent verification*, *ensure long-term compliance and enforcement*) depend on achieving low marginal costs of measurement to make independent, long-term verification feasible. The difference in emphasis arises because we consider auditability to be the foundation of a credible market.

The key question is whether all four criteria can be met *at scale*. We argue that recent developments in remote sensing, proximal sensing, citizen science, novel community data, and deep-learning species distribution models make this possibility newly realistic (Figure 1). These tools do not yet solve all challenges, especially those of causal attribution and compressing high-dimensional ecological information into parsimonious metrics. But they do open a credible path away from biodiversity governance based on coarse, weakly verifiable proxies and towards systems grounded in scalable, auditable ecological outcomes. In the remainder of this Roadmap, we assess how close current methods are to that goal and identify the scientific, political, and institutional steps needed to reach it.

The potential solution: Scalable Biodiversity Measurement

The difficulty in designing metrics that satisfy all four criteria arises because the same information is asked to serve two distinct purposes. Scientific understanding demands rich, high-dimensional data, whereas management effectiveness requires signals that are simple, cheap, but informative enough to support decisions, reporting, and assurance(Layman et al., 2024). Alleviating the trade-off entails applying technology (= methods that ‘expand what is feasible’) so that complex biodiversity observations can be transformed into auditable metrics without severing the link between them.

We argue that the necessary technology now exists in the combination of (1) remote and proximal sensing, (2) citizen science databases and novel community data, and (3) deep learning(Bush et al., 2023; Hartig et al., 2024; Jetz et al., 2019; Randin et al., 2020; Sutherland et al., 2026; Tosa et al., 2021; White et al., 2021). Combined, these technological advances enable **scalable biodiversity measurement** (Figure 1), by which we mean the mapping of biodiversity at fine resolution, across large areas, at costs compatible with frequent auditing.

At the centre of this capability are species distribution models (SDMs), which predict where species occur or, where data permit, their abundances. SDMs can in turn support a full range of conservation decisions, from identifying critical habitat and selecting reserve networks to strengthening the effectiveness and credibility of the policy and financial instruments on which restoration and conservation depend(Affinito et al., 2024; Antonelli et al., 2023; Beery et al., 2021; Bush et al., 2017; Carraro et al., 2020; Cavender-Bares et al., 2022; Giakoumi et al., 2025; Gonzalez et al., 2023; Griffith et al., 2026; Guisan et al., 2013; Hartig et al., 2024; Jetz et al., 2019; Kass et al., 2024; Mandle et al., 2024; Neubert et al., 2025; Planet et al., 2024; Pollock et al., 2020, 2025; Randin et al., 2020; Swinfield et al., 2024; Tosa et al., 2021; White et al., 2023; Xu et al., 2023).

However, SDMs have historically fallen short of their potential at local and landscape scales(Domisch et al., 2019). Most SDMs, which we term *classical SDMs*, have used coarse-resolution, infrequently updated environmental predictors like maps of climate, soil, and land cover(Mateo et al., 2019), which do not match the fine-grained, up-to-date needs of conservation decision-making(Araújo et al., 2019; Guisan et al., 2013; Mateo et al., 2019). Consequently, a review in the early 2010s deemed <1% of classical SDMs relevant for local and regional management decisions(Guisan et al., 2013).

This situation is changing because ecology is undergoing three parallel technological revolutions. First, *environmental sensing* is being scaled up via remote and proximal sensing systems, which provide temporally and spatially granular measurements across multiple spectral channels and other modalities in almost real-time, with some offering global coverage and others offering rapid, low-cost, and high-resolution measurements at local scale(Bush et al., 2017; Geller et al., 2023; He et al., 2015; Randin et al., 2020). Second, *species sensing* is being scaled up via citizen-science databases(iNaturalist, 2026; Kelling et al., 2012), compilations of national surveys and expert knowledge(Chytrý et al., 2015; GBIF, 2026), and novel community data—a term for high-throughput, standardised point samples of biodiversity made possible by coupling DNA sequencing, audio recorders, and cameras with species identification algorithms(Hartig et al., 2024). Third, modelling is being transformed by deep learning, which can identify complex, nonlinear, and generalisable linkages between environmental measurements and species observations when sufficient and appropriate training data is available, an approach we will refer to as DL-SDMs (deep-learning-based SDMs)(Pollock et al., 2025; Xu et al., 2023).

Together, these technological advances make scalable biodiversity measurement plausible in a way that was not previously possible: large numbers of species distributions can be mapped at fine resolution across tens of thousands of square kilometres, with frequent updates driven by changes in remotely sensed variables and targeted local measurements. The long-standing gap between ecologically meaningful biodiversity data and governance-ready biodiversity metrics is beginning to narrow.

But can DL-SDMs satisfy the four criteria of credible biodiversity metrics? Our assessment is favourable but mixed. Criterion 1, low marginal cost, is achievable now, provided that capital is invested in large, standardised training datasets for each biome–region combination. Criterion 2, low dispersion, can likely be met at high-priority sites by adding proximal sensing of environmental conditions and partial species compositions, albeit with an increase in marginal cost(Coverdale et al., 2025; Godfree & Knerr, 2025; Huynh et al., 2024; Müller et al., 2023; Schneck et al., 2023; Shi et al., 2026).

By contrast, Criterion 3, sufficient information content, is only partially satisfied at present; DL-SDMs can identify areas of higher and lower conservation value, but robust causal attribution requires further methodological advances. Criterion 4, parsimony, presents a different challenge because it requires compressing high-dimensional SDM outputs into one or a few socially legitimate metrics for governance and assurance. Science can propose compression algorithms and define core factors (e.g. composition, function, trophic complexity, disturbance, connectivity, and site irreplaceability)(Bush et al., 2023; Czúcz et al., 2021; Nicholson Thomas et al., 2025; Perino et al., 2019), but deciding what information must be retained and what can be lost is ultimately a socio-political choice.

In the following sections, we review DL-SDMs, assess their readiness to audit claims across the four steps of the Mitigation Hierarchy, and outline a roadmap for accelerating the deployment of scalable biodiversity measurement.

Deep-Learning-based Species Distribution Models

Like SDMs generally, a DL-SDM takes as input several environmental measurements and species observations in each of a set of locations (the training dataset) and learns to predict the probability of presence of each species continuously across the landscape. DL-SDMs are more complex than conventional and machine-learning

SDMs, which has been found to result in better predictive performance for data-rich species(Kellenberger et al., 2026) (Table S2).

DL-SDMs can also use deep learning to generate the environmental measurements themselves. For instance, a convolutional neural network (CNN) can compress a remote-sensing image centred over a species observation into a lower-dimensional vector (known as an embedding) that encodes both gross and subtle environmental features, including microclimate, connectivity, structure, vegetation composition, soil, hydrology, terrain, and landscape context. When time series are available, a CNN can also encode seasonal dynamics such as vegetation phenology. The embeddings are then used as predictors of species presence/absence. This ability to infer rich environmental structure from imagery is one reason that DL-SDMs are expected to perform better than classical approaches, which can only use tabular predictors(Gillespie et al., 2024).

A further advantage is that DL-SDMs can be trained on a large number of locations and on hundreds up to thousands of species simultaneously, which allows the models to learn general environment–occurrence relationships, improving predictions for rare and data-deficient species relative to models trained on each species independently(Deneu et al., 2021; Dinnage, 2024; Estopinan et al., 2022; Kellenberger et al., 2026; Pollock et al., 2025). The visual information from imagery is usually combined with tabular climate, soil, and topographic data into so-called multimodal models, which have been found to outperform unimodal versions(Dollinger et al., 2024; Hu et al., 2025; Tiel et al., 2025). We describe DL-SDMs in more detail in *Supplementary Information 3* and *Table S2*.

The proximate output of a DL-SDM is a stack of species distribution maps, one per species, at the spatial resolution of the input imagery — typically 10–50 m for satellite-based models and 0.2–1 m for airborne imagery. One of the largest DL-SDMs to date mapped 5,558 plant species across 5.5 billion grid cells in Europe(Leblanc et al., 2025). The individual species maps can be aggregated into summary metrics such as species richness and composition(Hu et al., 2025), habitat classes(Leblanc et al., 2025), and site irreplaceability(Li et al., 2024). Other studies have used supervised classification to learn to predict summary metrics directly, like species richness(Baggström et al., 2025), forest types(Picard et al., 2024), seagrass extent(Giménez-Romero et al., 2024), and protected-versus-anthropogenic land cover (Stomberg et al., 2023). Either way, once trained, these models can generate predictions from freely available satellite imagery at low cost per additional square kilometre, which is why DL-SDMs are especially promising with respect to Criterion 1, low marginal cost.

Currently, general conclusions about the relative and absolute predictive performances of DL-SDMs are challenging to make because studies differ substantially in the species and predictor data and evaluation metrics used (Table S2). In particular, most DL-SDM studies have used presence-only datasets, which are attractive because of their sizes but for which no single evaluation metric is foolproof. However, in a deliberately apples-to-apples comparison, Kellenberger et al.(2026) compared two deep learning (multilayer perceptron and CNN) to two classical (Random Forest and Maxent) SDMs for their ability to use coarse, tabular predictors (climate, productivity, topography, and soil) to model presence-only data for 2299 terrestrial vertebrate and invertebrate species. Across all species, the classical models achieved slightly higher mean AUC values (MLP: 0.78, CNN: 0.76, Maxent: 0.81, RF: 0.81), but for data-rich species (>200 observations), the DL-SDMs performed absolutely and relatively better.

Two studies that also allowed DL-SDMs to use remote-sensing covariates, which classical SDMs cannot use directly, have reported higher performance values for the DL-SDMs. Davis et al.(2023) found that their deep-learning implementation of a joint species distribution model "consistently outperformed" a Bayesian JSDM across eight datasets and six evaluation metrics, and Gillespie et al.(2024) reported that their DL-SDM achieved a median AUC of 0.95 versus 0.88 for Maxent and Random Forest.

We conclude that it is reasonable to expect DL-SDMs to achieve reliably high predictive performance for data-rich species. But for data-poor species, which include many threatened species, we should conservatively assume that DL-SDMs in their current form do not outperform classical approaches. Also, published DL-SDMs (Table S2) almost all model birds, plants, or vertebrates in Europe and North America, so we have little data on how well DL-SDMs perform for other taxa and biomes. (See Table S2 for additional studies and performance metrics.) Finally, our focus is on terrestrial ecosystems (including ponds and lakes). River networks(Altermatt et al., 2025) and marine ecosystems(Peng et al., 2026) will require modified approaches.

Extensions. – For data-sparse species, model extensions are being developed to improve performance with little effect on marginal costs. These extensions include transfer learning from large ecological and remote-sensing imagery datasets(Bourel et al., 2025; Brown et al., 2025; Kellenberger et al., 2026; Rolf et al., 2021; Stewart et al., 2023) and accounting for class imbalance(Gillespie et al., 2024; Kellenberger et al., 2026; Zbinden et al., 2024) (= high proportions of species absences) and observation error(Davis et al., 2023; Seo et al., 2021).

Other extensions improve predictive performance but increase marginal costs. Proximal sensors — e.g. passive acoustic recorders, cameras, and airborne lidar — provide fine-grained environmental measurements not visible to satellites(Baggström et al., 2025; Haucke et al., 2025; Li et al., 2024), and independent observations of some co-occurring species can be used to condition predictions of the other species at the same sites, exploiting species correlations learned during model training that carry information about unmeasured environmental features, community assembly rules, and biotic interactions(Abdelwahed et al., 2024, 2026). Proximal sensing and conditional prediction trade increases in marginal cost (Criterion 1) for gains in information content and statistical power (Criteria 2 and 3) and are therefore more justifiable when key pressures are invisible to remote sensors, such as toxins and hunting, and when decision stakes are high, such as where legally protected species might be present.

Capability gaps. – Finally, important capabilities remain underdeveloped. Species share phylogenetic and trait similarities, and models that exploit these similarities during training should improve predictions for rare species. Explicit phylogenetic or trait-informed sharing has been proposed for DL-SDMs(Harrell et al., 2025) but not yet demonstrated to improve rare-species predictions, though a Bayesian joint SDM has used phylogenetic relatedness to regularise the parameters of rare arthropod species(Ovaskainen et al., 2025).

Most critically for the governance applications motivating this paper, current DL-SDMs lack formal causal-attribution machinery(Schrodt et al., 2025). Gillespie et al.(2024) offer a suggestive result: their DL-SDM correctly predicted temporal community turnover following wildfire and spatial turnover across ecotonal and successional gradients, but the authors caution that the predictions remain correlational. Supervised driver classification can identify the likely driver of a detected change. For example, Sims et al.(2025) human-labelled 7,000 images by most likely deforestation driver (agriculture, mining, etc.) to train a model that mapped proximate causes of forest loss globally, but this approach cannot infer what would have happened absent the driver. For that, counterfactuals are required. Here, the large spatial extents of DL-SDMs are an asset, because they facilitate the identification of covariate-matched counterfactual sites(Andam et al., 2008; Carter et al., 2026; Ren et al., 2015; Swinfield et al., 2024).

The deeper problem is that DL-SDMs optimised for predictive accuracy will exploit non-causal, merely correlational relationships among collinear features(Shmueli, 2010). That weakens their usefulness for predicting the outcomes of interventions and conflicts with regulator and auditor expectations that models should ‘predict the right things for the right reasons’(Coeckelbergh, 2020). A potential but currently remote possibility would be to constrain model features using a causal graph of known driver–biodiversity relationships(Schrodt et al., 2025; Zhao & Hastie, 2021). Explainable-AI diagnostics(Molnar, 2022; Ryo et al., 2021; Zhou & Ryo, 2025) would be used to verify *post hoc* that causal learning has succeeded. However, until and if such methods develop, the inability of DL-SDMs (and SDMs generally) to attribute specific causes to model-predicted biodiversity changes remains the most important capability gap for compliance and liability applications, where regulators and buyers of biodiversity credits need causal evidence (Criterion 3), not just correlational maps.

Readiness of scalable biodiversity measurement for the Mitigation Hierarchy

We assess the extent to which current DL-SDMs can improve the auditing of claims across the four steps of the Mitigation Hierarchy (Avoid, Minimise, Remediate, Offset)(Arlidge et al., 2018) (Figure 1). These steps capture the full range of actions and impacts that are required to be disclosed in sustainability reporting frameworks(Bush et al., 2017; Cavender-Bares et al., 2022; CSRD, 2022; Deutz et al., 2020; Kass et al., 2024; Swinfield et al., 2024; White et al., 2023).

Avoidance actions and Site irreplaceability

Avoidance actions prevent biodiversity loss by eliminating impacts at the outset, for example by siting infrastructure away from sensitive habitats or excluding high-conservation-value sites from logging. The credibility of avoidance claims therefore depends on the availability of spatially detailed maps of conservation value covering the area of interest. One candidate method is *systematic conservation planning*(Ball et al., 2009; Giakoumi et al., 2025; Kukkala & Moilanen, 2013; Moilanen et al., 2009; Silvestro et al., 2022) (aka spatial conservation prioritisation or SCP), which analyses overlapping species-range maps to identify the parts of a landscape that should be prioritised for conservation, given constraints such as budgets, threat levels, connectivity, and range-protection targets for different species.

A key output of an SCP analysis is a ranking of sites by their contribution to the “efficient achievement of conservation objectives”(Baisero et al., 2022; Bush et al., 2023; Kukkala & Moilanen, 2013), known as *site irreplaceability*. Sites with higher irreplaceability values are characterised by large numbers of species with small ranges and/or high range-protection targets(Li et al., 2024). By contrast, a site with a low irreplaceability value means that the site’s loss would still allow the protection target of every species in that site to be achieved elsewhere in the landscape. Physical features, such as mapped habitat classes(Si-Moussi et al., 2025), can also be input into SCP algorithms and ranked for conservation.

In this framework, avoidance means locating operations in areas of low irreplaceability. By linking species to trait and species-interaction databases, actors could avoid sites that are important for ecosystem services, such as refugia in agricultural landscapes that support pollination and natural pest control(Bush et al., 2017; Giakoumi et al., 2025; He et al., 2015; Kass et al., 2024). This integration of systematic conservation planning and SDMs—termed *sideways biodiversity modelling* by Pollock et al.(2020)—also allows conservation priorities to be updated as new information becomes available(Randin et al., 2020; Silvestro et al., 2022).

Minimisation actions

Minimisation actions reduce the magnitude, extent, or duration of biodiversity impacts that cannot be fully avoided. *Spatial minimisation* involves shifting project footprints toward areas of lower conservation value and can therefore be defined and quantified in the same way as Avoidance. *Operational minimisation*, however, is much harder to audit(Bull et al., 2025). Here, the action is adjusting the timing, intensity, or methods of activities to reduce harm. Verifying such claims requires demonstrating that species distributions contracted less at the impacted site relative to a counterfactual in which the minimisation measure was not applied. In many cases, this standard will be impossible to meet, and auditing may need to focus instead on whether the minimisation actions were implemented as claimed, supported by evidence from comparable cases showing that such actions are likely to deliver the expected benefits(White et al., 2023). Future DL-SDMs, coupled with proximal sensor data, could contribute to that evidentiary base, but only if methods of causal attribution are developed (*Supplementary Information 3*).

Remediation actions

Remediation actions aim to recover biodiversity at the impacted site, for example by replanting vegetation or re-establishing soil structure and hydrological processes. Auditing remediation requires detecting localised expansions in species distributions into restored areas and attributing these changes to the intervention itself. This is challenging because the relevant landscape modifications may be only weakly or not at all detected in remote-sensing predictors (e.g. suppression of poaching). Antonelli et al.(2023) make the complementary case that SDMs updated frequently using remote-sensing data can act as an early-warning system, identifying land-use changes that increase risks to particular sites or species and prompting timely updates to spatial priorities. Given the expected stochasticity of community assembly during remediation, and the novelty of post-remediation signals relative to the original training data, reliable verification is likely to require proximal sensing—such as bioacoustics, camera traps, video, or eDNA—to improve local species inference. These measurements must be repeated frequently enough to detect change from remediation. Proximal sensing thus increases the marginal cost of biodiversity auditing but could be justified in medium and high-impact scenarios.

Offsetting actions

Offsetting actions compensate for biodiversity losses by generating equivalent gains away from the impacted site, for example through habitat remediation or by conserving sites under threat. For *remediation offsets*, auditing must again rely on a combination of remote and proximal sensing to detect biodiversity recovery over time. For *conservation offsets*, the central challenge is to demonstrate additionality: demonstrating that protection delivers greater biodiversity outcomes than would have occurred in the absence of action. DL-SDMs can contribute in two ways. First, they can identify the sites where protection will produce the greatest gains, which Kujala et al.(2015) have shown is more efficient than like-for-like offsets and avoids the problem of needing to establish offset equivalence. Second, by mapping biodiversity across large, continuous areas, they can identify counterfactual sites against which outcomes can be assessed(Swinfield et al., 2024). DL-SDMs thus enable No Net Loss claims to be grounded in direct measurement of species distributions.

In summary, our assessment suggests that the components needed to audit Avoidance, Spatial Minimisation, and Conservation Offsets now exist, although the supporting infrastructure and training datasets must still be built. Moreover, when both the reporting actor and auditor rely on the same map for avoidance, spatial minimisation, or conservation offsets, low dispersion (Criterion 2) is effectively achieved.

Operational minimisation and remediation remain immature, requiring further advances in causal attribution. Also, sometimes a higher evidentiary standard is desired, crediting an action only if it results in higher species population viabilities(Fletcher et al., 2025; Fordham et al., 2013; Guisan et al., 2013); this will raise data and modelling requirements, and thus marginal costs.

A Roadmap to Credible Biodiversity Metrics

We conclude with a three-step roadmap for converting scalable measurement into credible biodiversity metrics: (1) capital investment in training data, (2) a transparent political process to agree on dimension reduction algorithms, and (3) an integration of biodiversity science with mechanism-design economics, so that policies and financial instruments can robustly deploy these new technological capabilities (Figure 1).

Capital investment to create large training datasets in every biome-region combination

Nearly all deep learning SDMs have mined existing citizen-science and national databases (*Supplementary Information 3, Table S2*). While these provide massive sample sizes, they are limited to narrow taxonomic slices, dominated by presence-only data, and geographically biased toward Western Europe and North America. The field must now integrate data analysis with large-scale data generation because the main obstacle to achieving high predictive performance in DL-SDMs is the lack of large, training datasets.

Thus, our first recommendation is for capital investments to generate large, high-quality, and taxonomically broad presence–absence datasets for every biome–region combination globally(Meyer & Pebesma, 2022; Nyström et al., 2026; Sutherland et al., 2026). Initial sampling efforts should prioritise areas of high conservation urgency and aim to generate biodiversity data over at least tens of thousands of square kilometres.

This minimum spatial extent is important for both ecological and economic reasons. Ecologically, it is the scale required to capture landscape-scale processes—such as dispersal, migration corridors, and source–sink dynamics—that underpin long-term population viability and biodiversity persistence. Economically, this extent aligns more closely with the operational footprints of sub-national administrative regions and corporate supply chains, providing the continuous surface needed for systematic conservation planning and the verification of avoidance and conservation-offset claims.

Ideally, these datasets would be updated regularly to produce time series(Hartig et al., 2024), which would increase predictive performance(Anselmetto et al., 2025) and increase the ability of DL-SDMs to attribute specific causes to model-predicted biodiversity changes (Criterion 3).

To ensure transparency and scalability, data and models should be built using standardised sampling protocols and hosted on public registers —similar to eBird— so that multiple organisations and even individuals can continuously contribute data, build models, and record mitigation actions(Kim et al., 2025; Kujala et al., 2022; Miller et al., 2025; Sutherland et al., 2026). Furthermore, high resolution habitat maps(Si-Moussi et al., 2025) and/or an active learning approach should be used to direct sampling toward the most informative sites(Ryan et al., 2016; Xu et al., 2023). A public-ownership strategy of this kind would reduce the data-fragmentation, low-legitimacy, and high-acquisition-cost problems likely to arise if biodiversity measurement becomes dominated by proprietary models developed by private technology firms(ERM et al., 2024; Planet et al., 2024).

A transparent political process to agree on a parsimonious set of metrics

In our view, the most difficult challenge is reducing the complex outputs of biodiversity measurement to just a few metrics that are understandable, auditable, and mitigate against selective disclosures(Dietz et al., 2003; Layman et al., 2024). This is the challenge and promise of parsimony. The objective is to construct what are known as *bottom-up* metrics(Hawkins et al., 2024), which are aggregated from primary measurements of individual biodiversity components, i.e. individual species distributions, without severing the link between the reported value and the underlying ecological data. Only bottom-up metrics enable users to “calculate actual, realised footprints on the ground and track outcomes of management decision making”(Bedford et al., 2026; Burgess et al., 2024). However, while many bottom-up metrics have been proposed(Burgess et al., 2024; Fletcher et al., 2025; Nature Positive Initiative, 2026; Perino et al., 2019; Schneck et al., 2023; Wallacea Trust, 2023; Westoby et al., 2026), they suffer from high marginal costs (failing Criterion 1) and probably also from high dispersion (failing Criterion 2). Conversely, *top-down* measures of biodiversity intactness like Mean Species Abundance, which are low marginal cost because global maps are available(Bulle et al., 2019; Schipper et al., 2020), have been deemed as unsuitable for disclosures and assurance, due to high dispersion and error, and to low information content, thus failing criteria 2 and 3(Burgess et al., 2024; Hawkins et al., 2024; Mendez Angarita et al., 2025; Meyer & Pebesma, 2022; Miller et al., 2025; Nicholson Thomas et al., 2025; Silvestro et al., 2025).

Thus, our second recommendation is for a transparent political process to agree on dimension-reduction algorithms(Miller et al., 2025). Science can propose algorithms, but metrics are value-laden(Westoby et al., 2026), and thus stakeholders must decide together on how much information loss is acceptable. In this respect, dimension reduction is an application of what is known broadly as multi-criteria decision-making and multi-attribute utility, for which there is a huge literature(Adem Esmail & Geneletti, 2018; Ananda & Herath, 2009; Keeney, 1977; Neubert et al., 2025) that we will not delve into here.

However, following our principal-agent model (*Supplementary Information 2*), we take the pragmatic position that biodiversity metrics need not capture everything. They need only retain enough information, with sufficiently low dispersion, to satisfy informed regulators, auditors, consumers, and financial institutions that relevant impacts have been correctly assessed for the decision context (e.g. consumer certification, environmental impact assessment, payments for ecosystem services, ethical finance, biodiversity credits issuance). Domain experts, especially ecologists, must inform these decisions, but ultimately, the decisions are social and political rather than purely scientific. Information loss is the cost paid to gain effective management and policies(Needham et al., 2019).

At the same time, even a single metric can carry substantial information if well designed. Site irreplaceability and other information-rich metrics being proposed(Ferrier, 2026; Fletcher et al., 2025; Kukkala & Moilanen, 2017; Maczik et al., 2024; Rossberg et al., 2024; Silvestro et al., 2025) can integrate goals for species composition, ecosystem functioning, habitat structure, and connectivity.

Integration of biodiversity science with mechanism-design economics

The deep challenge is that biodiversity governance operates under persistent asymmetric information. Regulators, investors, and the public cannot directly observe most ecological impacts, and this limits society's ability to steer the actions of firms, governments, and other actors. That constraint forces us to accept tradeoffs in the design of nature-positive policies and markets and cannot be treated as a problem to be solved later, because policies and instruments that do not build in auditability will fail in practice. Fortunately, we are moving to a world of scalable biodiversity measurement, reducing its marginal cost while increasing the quantity and quality of ecological information.

Yet technology and policy must develop together because new information systems alone do not automatically generate well-functioning institutions(Kim et al., 2025; Wunder et al., 2025). Our third recommendation is therefore to integrate biodiversity science with *mechanism design*, the branch of economics concerned with designing institutions that align individual incentives with social goals under asymmetric information (*Supplementary Information 2).* For a given technology level, economists can design biodiversity credit markets that are robust to model uncertainty and ecological stochasticity; identify when the added costs of proximal sensing, conditional prediction, replicated sampling, or additional modelling are justified by gains in statistical precision and information content; evaluate the market consequences of different dimension-reduction algorithms used to construct metrics; and establish the evidence rules required to trigger payments upon achieving specific Mitigation Hierarchy targets. More broadly, they can identify which scientific and technological investments would yield the highest returns across the four criteria for credible biodiversity metrics.

Concluding section

Auditability cannot be added on after the fact but must be designed into policy and financial instruments from the start. In ecology and conservation, the received wisdom when building biodiversity credits, net-gain rules, or disclosure standards has been to specify a complete ecological metric and worry about checking it afterwards. Our argument is that this is backwards. When one side can hide what it knows, a rich metric that no one can trust is worth less than a coarse one that everyone can.

This likely means giving something up: an auditable metric will miss some real biodiversity. However, we are fortunate that recent technological advances in modelling and in species and environmental sensing should now let us reduce the loss. Building in auditability from the start is what turns biodiversity measurement into credible metrics, and credible metrics into policy that both scales and works.

Acknowledgments

We thank Florian Hartig for extensive advice and comments, especially on the importance of proximal sensing, dimension reduction, and causal attribution.

Funding statement

DWY was supported by the National Key Research and Development Program of China (2022YFC2602500), the Strategic Priority Research Program of Chinese Academy of Sciences, Grant (XDA20050202), the National Natural Science Foundation of China (41661144002), and the Yunnan Revitalization Talent Support Program: High-end Foreign Expert Project. VDP was supported by a Columbia University President's Global Innovation Fund grant, a grant from the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project number PN-IV-P1-PCE-2023-1119 (Harmonia), within PNCDI IV and by the Chinese Academy of Science President's International Fellowship Initiative (2024VBB0002). GM was supported by UNIPD-Fondazione

Cariparo Starting Package. TB, TG and HOØ was supported by Norwegian research council (grant 344158-Biodiversity mapping of forest from above).

Competing interests

DWY is a co-founder of the commercial eDNA service company NatureMetrics but is not involved in its operations or management. The other authors have no competing interests.

Use of generative AI

We created a Claude skill (github.com/dougwyu/claude-zotero-skills) to copyedit the manuscript and to audit 'citation faithfulness' by comparing the semantic contents of cited papers with the semantic content of their citing sentences.

References

Abdelwahed, H. R., Teng, M., & Rolnick, D. (2024). *Predicting Species Occurrence Patterns from Partial Observations* (arXiv:2403.18028). arXiv. https://doi.org/10.48550/arXiv.2403.18028

Abdelwahed, H. R., Teng, M., Zbinden, R., Pollock, L., Larochelle, H., Tuia, D., & Rolnick, D. (2026). CISO: Species distribution modelling Conditioned on Incomplete Species Observations. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.70238

Adem Esmail, B., & Geneletti, D. (2018). Multi-criteria decision analysis for nature conservation: A review of 20 years of applications. *Methods in Ecology and Evolution*, *9*(1), 42–53. https://doi.org/10.1111/2041-210X.12899

Affinito, F., Williams, J. M., Campbell, J. E., Londono, M. C., & Gonzalez, A. (2024). Progress in developing and operationalizing the Monitoring Framework of the Global Biodiversity Framework. *Nature Ecology & Evolution*, *8*(12), 2163–2171. https://doi.org/10.1038/s41559-024-02566-7

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*, 488–500. https://doi.org/10.2307/1879431

Altermatt, F., Couton, M., Carraro, L., Keck, F., Lawson-Handley, L., Leese, F., Zhang, X., Zhang, Y., & Blackman, R. C. (2025). Utilizing aquatic environmental DNA to address global biodiversity targets. *Nature Reviews Biodiversity*. https://doi.org/10.1038/s44358-025-00044-x

Ananda, J., & Herath, G. (2009). A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics*, *68*(10), 2535–2548. https://doi.org/10.1016/j.ecolecon.2009.05.010

Andam, K. S., Ferraro, P. J., Pfaff, A., Sanchez-Azofeifa, G. A., & Robalino, J. A. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences*, *105*(42), 16089–16094. https://doi.org/10.1073/pnas.0800437105

Anselmetto, N., Garbarino, M., Weldy, M. J., Bell, D. M., Daly, C., Epps, C. W., Ferrari, N., Kim, H., LaManna, J. A., Lesmeister, D. B., Penaluna, B. E., Schulze, M., Sutton, M., Tosa, M. I., & Betts, M. G. (2025). Leveraging long-term data to improve biodiversity monitoring with species distribution models. *Journal of Applied Ecology*, *62*(11), 2914–2929. https://doi.org/10.1111/1365-2664.70177

Antonelli, A., Dhanjal-Adams, K. L., & Silvestro, D. (2023). Integrating machine learning, remote sensing and citizen science to create an early warning system for biodiversity. *PLANTS, PEOPLE, PLANET*, *5*(3), 307–316. https://doi.org/10.1002/ppp3.10337

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Arlidge, W. N. S., Bull, J. W., Addison, P. F. E., Burgass, M. J., Gianuca, D., Gorham, T. M., Jacob, C., Shumway, N., Sinclair, S. P., Watson, J. E. M., Wilcox, C., & Milner-Gulland, E. J. (2018). A Global Mitigation Hierarchy for Nature Conservation. *BioScience*, *68*(5), 336–347. https://doi.org/10.1093/biosci/biy029

Baggström, A., Goodsell, R., Van Dijk, L., Iwaszkiewicz-Eggebrecht, E., Miraldo, A., Tack, A. J. M., & Andermann, T. (2025). The utility of combining deep learning with metabarcoding to model biodiversity dynamics at a national scale. *Ecological Informatics*, *90*, 103318. https://doi.org/10.1016/j.ecoinf.2025.103318

Baisero, D., Schuster, R., & Plumptre, A. J. (2022). Redefining and mapping global irreplaceability. *Conservation Biology*, *36*(2), e13806. https://doi.org/10.1111/cobi.13806

Ball, I. R., Possingham, H. P., & Watts, M. (2009). Marxan and relatives: Software for spatial conservation prioritisation. In A. Moilanen, K. A. Wilson, & H. P. Possingham (Eds.), *Spatial conservation prioritisation: Quantitative methods and computational tools* (pp. 185–195). (Original work published Oxford University Press)

Bedford, J., Houdet, J. R. A., Berger, J., Grigg, A., Harrison, M., Calhoun, E., & Brooks, S. (2026). Measurement approaches for corporate impacts on ecosystem condition: Current landscape and future priorities. *Methods in Ecology and Evolution*, *17*(5), 1336–1344. https://doi.org/10.1111/2041-210x.70182

Beery, S., Cole, E., Parker, J., Perona, P., & Winner, K. (2021). Species Distribution Modeling for Machine Learning Practitioners: A Review. *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS '21*, 329–348. https://doi.org/10.1145/3460112.3471966

Blanco-Zaitegi, G., Álvarez Etxeberria, I., & Moneva, J. M. (2024). Impression management of biodiversity reporting in the energy and utilities sectors: An assessment of transparency in the disclosure of negative events. *Journal of Behavioral and Experimental Finance*, *42*, 100942. https://doi.org/10.1016/j.jbef.2024.100942

Boiral, O., & Heras-Saizarbitoria, I. (2020). Sustainability reporting assurance: Creating stakeholder accountability through hyperreality? *Journal of Cleaner Production*, *243*, 118596. https://doi.org/10.1016/j.jclepro.2019.118596

Bourel, B., Joly, A., Servajean, M., Bettinger, S., Sanabria-Fernández, J. A., & Mouillot, D. (2025). From Presence-Only to Abundance Species Distribution Models Using Transfer Learning. *Ecology Letters*, *28*(7), e70177. https://doi.org/10.1111/ele.70177

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., & Kohli, P. (2025). *AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data* (arXiv:2507.22291). arXiv. https://doi.org/10.48550/arXiv.2507.22291

Bull, J. W., Taylor, I., De Valença, A., IJspeert, R., Van Erve, B., Modernel, P., & Poore, J. A. C. (2025). Towards positive net outcomes for biodiversity, and developing safeguards to accompany headline biodiversity indicators. *Npj Biodiversity*, *4*(1), 31. https://doi.org/10.1038/s44185-025-00095-5

Bulle, C., Margni, M., Patouillard, L., Boulay, A.-M., Bourgault, G., De Bruille, V., Cao, V., Hauschild, M., Henderson, A., Humbert, S., Kashef-Haghighi, S., Kounina, A., Laurent, A., Levasseur, A., Liard, G., Rosenbaum, R. K., Roy, P.-O., Shaked, S., Fantke, P., & Jolliet, O. (2019). IMPACT World+: A globally regionalized life cycle impact assessment method. *The International Journal of Life Cycle Assessment*, *24*(9), 1653–1674. https://doi.org/10.1007/s11367-019-01583-0

Burgess, N. D., Ali, N., Bedford, J., Bhola, N., Brooks, S., Cierna, A., Correa, R., Harris, M., Hargey, A., Hughes, J., McDermott-Long, O., Miles, L., Ravilious, C., Rodrigues, A. R., Soesbergen, A. van, Sihvonen, H., Seager, A., Swindell, L., Vukelic, M., … Butchart, S. H. M. (2024). Global Metrics for Terrestrial Biodiversity. *Annual Review of Environment and Resources*, *49*(Volume 49, 2024), 673–709. https://doi.org/10.1146/annurev-environ-121522-045106

Bush, A., Simpson, K. H., & Hanley, N. (2023). Systematic nature positive markets. *Conservation Biology*, cobi.14216. https://doi.org/10.1111/cobi.14216

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C. A., Dawson, T. P., Emerson, B. C., Ferrier, S., Gilbert, M. T. P., Herold, M., Jones, L., Leendertz, F. H., Matthews, L., Millington, J. D. A., … Yu, D. W. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution*, *1*(7), 0176. https://doi.org/10.1038/s41559-017-0176

Carraro, L., Mächler, E., Wüthrich, R., & Altermatt, F. (2020). Environmental DNA allows upscaling spatial patterns of biodiversity in freshwater ecosystems. *Nature Communications*, *11*(1), 3585. https://doi.org/10.1038/s41467-020-17337-8

Carter, H., Thompson, B. S., Bull, J. W., Dickman, A., Jones, J. P. G., Shrikanth, S., & zu Ermgassen, S. O. S. E. (2026). Demystifying biodiversity finance. *Nature Reviews Biodiversity*. https://doi.org/10.1038/s44358-026-00155-z

Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., Fatoyinbo, L., Hurtt, G. C., Schimel, D., Townsend, P. A., Ustin, S. L., Wang, Z., & Wilson, A. M. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology & Evolution*, *6*, 506–519. https://doi.org/10.1038/s41559-022-01702-5

Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J. H. J., Aćić, S., Agrillo, E., Ambarlı, D., Angelini, P., Apostolova, I., Attorre, F., Berg, C., Bergmeier, E., Biurrun, I., Botta-Dukát, Z., Brisse, H., … Yamalov, S. (2015). European Vegetation Archive (EVA): An integrated database of European vegetation plots. *Applied Vegetation Science*, *19*(1), 173–180. https://doi.org/10.1111/avsc.12191

Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

Convention on Biological Diversity. (2022). *Decision 15/4: Kunming-Montreal Global Biodiversity Framework. CBD/COP/DEC/15/4*. https://www.cbd.int/article/cop15-final-text-kunming-montreal-gbf-221222

Coverdale, T. C., Boucher, P. B., Singh, J., & Davies, A. B. (2025). Quantifying aboveground herbaceous biomass in grassy ecosystems: A comparison of field and high-resolution UAV-LiDAR approaches. *Remote Sensing in Ecology and Conservation*, rse2.70023. https://doi.org/10.1002/rse2.70023

CSRD. (2022). DIRECTIVE (EU) 2022/2464 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting. *Official Journal of the European Union*. http://data.europa.eu/eli/dir/2022/2464/oj

Czúcz, B., Keith, H., Driver, A., Jackson, B., Nicholson, E., & Maes, J. (2021). A common typology for ecosystem characteristics and ecosystem condition variables. *One Ecosystem*, *6*, e58218. https://doi.org/10.3897/oneeco.6.e58218

Dahm, M., González, P., & Porteiro, N. (2018). The enforcement of mandatory disclosure rules. *Journal of Public Economics*, *167*, 21–32. https://doi.org/10.1016/j.jpubeco.2018.08.014

Davis, C. L., Bai, Y., Chen, D., Robinson, O., Ruiz‐Gutierrez, V., Gomes, C. P., & Fink, D. (2023). Deep learning with citizen science data enables estimation of species diversity and composition at continental extents. *Ecology*, *104*(12), e4175. https://doi.org/10.1002/ecy.4175

de Groen, W. P., Alcidi, C., Bruggeman, V., Rossotti, T., Beatriz Pozo, Hassan, E., Oliinyk, I., Gojsic, D., Andersson, F., Goekten, M., Korenblit, A., Nagtegaal, J., Cinova, D., Ellina, D., Gasperini, M., Akhvlediani, T., Righetti, E., Sproge, J., & Vona, L. (2022). *Cost-benefit analysis of the First Set of draft European Sustainability Reporting Standards: FINAL REPORT* (p. 136). CEPS and Milieu. https://www.ceps.eu/ceps-publications/cost-benefit-analysis-of-the-first-set-of-draft-european-sustainability-reporting-standards/

Deneu, B., Servajean, M., Bonnet, P., Botella, C., Munoz, F., & Joly, A. (2021). Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLOS Computational Biology*, *17*(4), e1008856. https://doi.org/10.1371/journal.pcbi.1008856

Deutz, A., Heal, G. M., Niu, E., Swanson, E., Townshend, T., Li, Z., Adams, A., Meghji, A., Sethi, S. A., & Tobin-de la Puente, J. (2020). *Financing Nature: Closing the global biodiversity financing gap. Full Report*. The Paulson Institute, The Nature Conservancy, and the Cornell Atkinson Center for Sustainability. https://www.paulsoninstitute.org/conservation/financing-nature-report/

Dietz, T., Ostrom, E., & Stern, P. C. (2003). The Struggle to Govern the Commons. *Science*, *302*(5652), 1907–1912. https://doi.org/10.1126/science.1091015

Dinnage, R. (2024). *NicheFlow: Towards a foundation model for Species Distribution Modelling*. BioRXiv. https://doi.org/10.1101/2024.10.15.618541

Dollinger, J., Brun, P., Sainte Fare Garnot, V., & Wegner, J. D. (2024). Sat-SINR: High-Resolution Species Distribution Models Through Satellite Imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-2–2024*, 41–48. ISPRS TC II Mid-term Symposium "The Role of Photogrammetry for a Sustainable World" - 11–14 June 2024, Las Vegas, Nevada, USA. https://doi.org/10.5194/isprs-annals-X-2-2024-41-2024

Domisch, S., Friedrichs, M., Hein, T., Borgwardt, F., Wetzig, A., Jähnig, S. C., & Langhans, S. D. (2019). Spatially explicit species distribution models: A missed opportunity in conservation planning? *Diversity and Distributions*, *25*(5), 758–769. https://doi.org/10.1111/ddi.12891

Ducros, A., & Steele, P. (2022). *Biocredits to finance nature and people: Emerging lessons*. IIED. iied.org/21216iied

Elliot, V., Jonäll, K., Paananen, M., Bebbington, J., & Michelon, G. (2024). Biodiversity reporting: Standardization, materiality, and assurance. *Current Opinion in Environmental Sustainability*, *68*, 101435. https://doi.org/10.1016/j.cosust.2024.101435

ERM, ERM Sustainability Institute, Salesforce, NatureMetrics, & Planet. (2024). *The Nature Tech Revolution: The Tools to Move from 'Do No Harm' to 'Nature-Positive.'* https://www.erm.com/globalassets/sustainability.com/reports/ermsi_nature_tech_alliance_2410_v7.pdf

Estopinan, J., Servajean, M., Bonnet, P., Munoz, F., & Joly, A. (2022). Deep Species Distribution Modeling From Sentinel-2 Image Time-Series: A Global Scale Analysis on the Orchid Family. *Frontiers in Plant Science*, *13*, 839327. https://doi.org/10.3389/fpls.2022.839327

Ferrier, S. (2026). A nature-positive world is more than the sum of its parts. *Methods in Ecology and Evolution*, *17*(5), 1405–1417. https://doi.org/10.1111/2041-210x.70153

Fletcher, R. J., Green, R. E., Bladon, E. K., Atkinson, P. W., Phalan, B. T., Williams, D., Visconti, P., & Balmford, A. (2025). Beyond Species Richness for Biological Conservation. *Conservation Letters*, *18*(4), e13124. https://doi.org/10.1111/conl.13124

Fordham, D. A., Mellin, C., Russell, B. D., Akçakaya, R. H., Bradshaw, C. J. A., Aiello-Lammens, M. E., Caley, J. M., Connell, S. D., Mayfield, S., Shepherd, S. A., & Brook, B. W. (2013). Population dynamics can be more important than physiological limits for determining range shifts under climate change. *Global Change Biology*, *19*(10), 3224–3237. https://doi.org/10.1111/gcb.12289

Forest Trends' Ecosystem Marketplace. (2024). *State of the Voluntary Carbon Market 2024.* Forest Trends Association. https://www.ecosystemmarketplace.com/publications/2024-state-of-the-voluntary-carbon-markets-sovcm/

GBIF. (2026, April 10). *GBIF | Global Biodiversity Information Facility*. GBIF | Global Biodiversity Information Facility. https://www.gbif.org/

Geller, G., Levick, S., Luque, S., Sayre, R., Brink, A., Coops, N., Durrieu, S., Ewald, J., Feilhauer, H., Feret, J.-B., Ferrier, S., Honzak, M., Lenco, D., Koltz, A., Lele, N., Lombardo, S., Mahecha, M., Meyer, C., Paganini, M., … Viciano, L. (2023). *Space-based Earth observation and ecosystem extent: Exploring opportunities* [CEOS Ecosystem Extent Task Team White Paper V1.1 28 November 2023]. CEOS. https://ceos.org/document_management/Publications/Publications-and-Key-Documents/Cross-Cutting/EETT%20White%20Paper%20V1.1%202023-12-04.pdf

Giakoumi, S., Richardson, A. J., Doxa, A., Moro, S., Andrello, M., Hanson, J. O., Hermoso, V., Mazor, T., McGowan, J., Kujala, H., Law, E., Álvarez-Romero, J. G., Magris, R. A., Gissi, E., Arafeh-Dalmau, N., Metaxas, A., Virtanen, E. A., Ban, N. C., Runya, R. M., … Katsanevakis, S. (2025). Advances in systematic conservation planning to meet global biodiversity goals. *Trends in Ecology & Evolution*, *40*(4), 395–410. https://doi.org/10.1016/j.tree.2024.12.002

Gillespie, L. E., Ruffley, M., & Exposito-Alonso, M. (2024). Deep learning models map rapid plant species changes from citizen science and remote sensing data. *Proceedings of the National Academy of Sciences*, *121*(37), e2318296121. https://doi.org/10.1073/pnas.2318296121

Giménez-Romero, À., Ferchichi, D., Moreno-Spiegelberg, P., Sintes, T., & Matías, M. A. (2024). *Mapping the distribution of seagrass meadows from space with deep convolutional neural networks* (2024.03.21.586047; p. 2024.03.21.586047). bioRxiv. https://doi.org/10.1101/2024.03.21.586047

Godfree, R., & Knerr, N. (2025). Rapid ecological data collection from 360-degree imagery using visualisation and immersive sampling in the R pannotator package. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.14472

Gonzalez, A., Chase, J. M., & O’Connor, M. I. (2023). A framework for the detection and attribution of biodiversity change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1881), 20220182. https://doi.org/10.1098/rstb.2022.0182

Griffith, J., Lord, J.-M., Catchen, M. D., Arce-Plata, M. I., Blanchet, F. G., Chandramohan, M., Diaz-Corzo, M. C., Gravel, D., Gutiérrez, C., Helfenstein, I. S., Hoban, S., Kass, J. M., Laikre, L., Larocque, G., Leigh, D. M., Leung, B., Mastretta-Yanes, A., Millette, K. L., Berbeo, M. A. M., … Gonzalez, A. (2026). BON in a Box: An Open and Collaborative Platform for Biodiversity Monitoring, Indicator Calculation, and Reporting. *BioScience*, *76*(4), 345–358. https://doi.org/10.1093/biosci/biaf189

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*(12), 1424–1435. https://doi.org/10.1111/ele.12189

Harrell, L., Kaeser-Chen, C., Ayan, B. K., Anderson, K., Conserva, M., Kleeman, E., Neumann, M., Overlan, M., Chapman, M., & Purves, D. (2025). *Heterogeneous graph neural networks for species distribution modeling* (arXiv:2503.11900). arXiv. https://doi.org/10.48550/arXiv.2503.11900

Hartig, F., Abrego, N., Bush, A., Chase, J. M., Guillera-Arroita, G., Leibold, M. A., Ovaskainen, O., Pellissier, L., Pichler, M., Poggiato, G., Pollock, L., Si-Moussi, S., Thuiller, W., Viana, D. S., Warton, D. I., Zurell, D., & Yu, D. W. (2024). Novel community data in ecology – properties and prospects. *Trends in Ecology & Evolution*, *39*(3), 280–293. https://doi.org/10.1016/j.tree.2023.09.017

Haucke, T., Harrell, L., Shen, Y., Klein, L., Rolnick, D., Gillespie, L., & Beery, S. (2025). *Deep Multi-modal Species Occupancy Modeling* (p. 2025.09.06.674602). bioRxiv. https://doi.org/10.1101/2025.09.06.674602

Hawkins, F., Beatty, C. R., Brooks, T. M., Church, R., Elliott, W., Kiss, E., Macfarlane, N. B. W., Pugliesi, J., Schipper, A. M., & Walsh, M. (2024). Bottom-up global biodiversity metrics needed for businesses to assess and manage their impact. *Conservation Biology*, *38*(2), e14183. https://doi.org/10.1111/cobi.14183

He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M., Schmidtlein, S., Turner, W., Wegmann, M., & Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, *1*(1), 4–18. https://doi.org/10.1002/rse2.7

Hu, Y., Si-Moussi, S., & Thuiller, W. (2025). Introduction to deep learning methods for multi-species predictions. *Methods in Ecology and Evolution*, *16*(1), 228–246. https://doi.org/10.1111/2041-210X.14466

Huynh, A. V., Gillespie, L. E., Lopez-Saucedo, J., Tang, C., Sikand, R., & Expósito-Alonso, M. (2024). *Contrastive ground-level image and remote sensing pre-training improves representation learning for natural world imagery* (2409.19439v1). arXiv. https://arxiv.org/abs/2409.19439v1

iNaturalist. (2026, April 10). *iNaturalist*. California Academy of Sciences and National Geographic Society. https://www.inaturalist.org/

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E.

(2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, *3*(4), 539–551. https://doi.org/10.1038/s41559-019-0826-1

Kass, J. M., Fukaya, K., Thuiller, W., & Mori, A. S. (2024). Biodiversity modeling advances will improve predictions of nature's contributions to people. *Trends in Ecology & Evolution*, *39*(4), 338–348. https://doi.org/10.1016/j.tree.2023.10.011

Kedward, K., zu Ermgassen, S., Ryan-Collins, J., & Wunder, S. (2023). Heavy reliance on private finance alone will not deliver conservation goals. *Nature Ecology & Evolution*, *7*(9), 1339–1342. https://doi.org/10.1038/s41559-023-02098-6

Keeney, R. L. (1977). The art of assessing multiattribute utility functions. *Organizational Behavior and Human Performance*, *19*(2), 267–310. https://doi.org/10.1016/0030-5073(77)90065-4

Kellenberger, B., Winner, K., & Jetz, W. (2026). The Performance and Potential of Deep Learning for Predicting Species Distributions. *Global Ecology and Biogeography*, *35*(1), e70184. https://doi.org/10.1111/geb.70184

Kelling, S., Gerbracht, J., Fink, D., Lagoze, C., Wong, W.-K., Yu, J. Y., Damoulas, T., & Gomes, C. (2012). eBird: A Human/Computer Learning Network for Biodiversity Conservation and Research. *Proceedings of the AAAI Conference on Artificial Intelligence*, *26*(2), 2229–2236. https://doi.org/10.1609/aaai.v26i2.18963

Kim, E. H., Dellecker, A., Field, R., Stephenson, P. J., & Schrodt, F. (2025). Towards high-integrity biodiversity credits: Balancing commensurability, ecological complexity and governance. *Proceedings of the Royal Society B: Biological Sciences*, *292*(2053), 20250990. https://doi.org/10.1098/rspb.2025.0990

Kujala, H., Maron, M., Kennedy, C. M., Evans, M. C., Bull, J. W., Wintle, B. A., Iftekhar, S. M., Selwood, K. E., Beissner, K., Osborn, D., & Gordon, A. (2022). Credible biodiversity offsetting needs public national registers to confirm no net loss. *One Earth*, *5*(6), 650–662. https://doi.org/10.1016/j.oneear.2022.05.011

Kujala, H., Whitehead, A. L., Morris, W. K., & Wintle, B. A. (2015). Towards strategic offsetting of biodiversity loss using spatial prioritization concepts and tools: A case study on mining impacts in Australia. *Biological Conservation*, *192*, 513–521. https://doi.org/10.1016/j.biocon.2015.08.017

Kukkala, A. S., & Moilanen, A. (2013). Core concepts of spatial prioritisation in systematic conservation planning. *Biological Reviews*, *88*(2), 443–464. https://doi.org/10.1111/brv.12008

Kukkala, A. S., & Moilanen, A. (2017). Ecosystem services and connectivity in spatial conservation prioritization. *Landscape Ecology*, *32*(1), 5–14. https://doi.org/10.1007/s10980-016-0446-y

Layman, H., Akçakaya, H. R., Irwin, A., zu Ermgassen, S., Addison, P., & Burgman, M. (2024). Short-term solutions to biodiversity conservation in portfolio construction: Forward-looking disclosure and classification-based metrics. *Business Strategy and the Environment*, *33*(3), 1778–1793. https://doi.org/10.1002/bse.3570

Leblanc, C., Picek, L., Deneu, B., Bonnet, P., Servajean, M., Palard, R., & Joly, A. (2025). Mapping Biodiversity at Very-High Resolution in Europe. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2340–2349. https://doi.org/10.1109/CVPRW67362.2025.00221

Li, Y., Devenish, C., Tosa, M. I., Luo, M., Bell, D. M., Lesmeister, D. B., Greenfield, P., Pichler, M., Levi, T., & Yu, D. W. (2024). Combining environmental DNA and remote sensing for efficient, fine-scale mapping of arthropod biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1904), 20230123. https://doi.org/10.1098/rstb.2023.0123

Maczik, D. M., Jansen, V. A. A., & Rossberg, A. G. (2024). *Evaluating Biodiversity Credits Using Metacommunity Modelling*. BioRXiv. https://doi.org/10.1101/2024.06.03.597228

Mandle, L., Shea, A., Soth, E., Goldstein, J. A., Wolny, S., Smith, J. R., Chaplin-Kramer, R., Sharp, R. P., & Patel, M. (2024). An open-source approach for measuring corporate impacts on ecosystem services and biodiversity. *Communications Earth & Environment*, *5*(1), 1–11. https://doi.org/10.1038/s43247-024-01797-7

Marshall, E., Southwell, D., Wintle, B. A., & Kujala, H. (2024). A global analysis reveals a collective gap in the transparency of offset policies and how biodiversity is measured. *Conservation Letters*, *17*(1), e12987. https://doi.org/10.1111/conl.12987

Marshall, E., Wintle, B. A., Southwell, D., & Kujala, H. (2020). What are we measuring? A review of metrics used to describe biodiversity in offsets exchanges. *Biological Conservation*, *241*, 108250. https://doi.org/10.1016/j.biocon.2019.108250

Mateo, R. G., Gastón, A., Aroca-Fernández, M. J., Broennimann, O., Guisan, A., Saura, S., & García-Viñas, J. I. (2019). Hierarchical species distribution models in support of vegetation conservation at the landscape scale. *Journal of Vegetation Science*, *30*(2), 386–396. https://doi.org/10.1111/jvs.12726

Mendez Angarita, V. Y., Larsen, P. B., Marcolin, L., & Di Marco, M. (2025). Reconciling Different Forms of Ecological Integrity to Aid the Kunming-Montreal Global Biodiversity Framework. *Conservation Letters*, *18*(1), e13088. https://doi.org/10.1111/conl.13088

Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*, *13*(1), 2208. https://doi.org/10.1038/s41467-022-29838-9

Miller, B. L., Lombardo, S., Rosenthal, A., O'Shea, T., Luers, A., Lavista-Ferres, J. M., Mandle, L., Ricketts, T. H., Polasky, S., & Brown, D. G. (2025). Corporate Biodiversity Reporting Can Be Scaled With AI and Earth Observation—But Will Miss the Point Without Guidance From Conservation Scientists. *Conservation Letters*, *18*(5), e13153. https://doi.org/10.1111/conl.13153

Moilanen, A., Wilson, K. A., & Possingham, H. P. (Eds.). (2009). *Spatial conservation prioritization: Quantitative methods and computational tools*. Oxford University Press.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (Second edition). Christoph Molnar.

Müller, J., Mitesser, O., Schaefer, H. M., Seibold, S., Busse, A., Kriegel, P., Rabl, D., Gelis, R., Arteaga, A., Freile, J., Leite, G. A., De Melo, T. N., LeBien, J. G., Campos-Cerqueira, M., Blüthgen, N., Tremlett, C. J., Böttger, D., Feldhaar, H., Grella, N., … Buřivalová, Z. (2023). Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests. *Nature Communications*, *14*(1), 6191. https://doi.org/10.1038/s41467-023-41693-w

Nature Positive Initiative. (2026, February). *Draft Measurement Guidance Executive Summary: State of Nature Metrics—February 2026*. Nature Positive Initiative. https://www.naturepositive.org/app/uploads/2026/02/Supporting-Information_Draft-Measurement-Guidance-Executive-Summary_Feb2026.pdf

Needham, K., de Vries, F. P., Armsworth, P. R., & Hanley, N. (2019). Designing markets for biodiversity offsets: Lessons from tradable pollution permits. *Journal of Applied Ecology*, *56*(6), 1429–1435. https://doi.org/10.1111/1365-2664.13372

Neubert, S., McGowan, J., Metcalfe, K., Hanson, J. O., Buenafe, K. C. V., Dabalà, A., Dunn, D. C., Everett, J. D., Possingham, H. P., Stelzenmüller, V., Estep, A., Ervin, J., & Richardson, A. J. (2025). Multiple-use spatial planning for sustainable development and conservation. *Trends in Ecology & Evolution*, *40*(11), 1126–1142. https://doi.org/10.1016/j.tree.2025.09.007

Nicholson Thomas, I., Bernardo, F., Cazacu, C., Černecký, J., Cortinovis, C., Czúcz, B., Duchkova, H., Kičić, M., Lange, S., Lecomte, X., Lim, Z. Y. V., Lotan, A., Mörtberg, U., Parretta, C., Rendón, P., Roche, P., Tanács, E., Vuletić, D., Zoumides, C., & Grêt-Regamey, A. (2025). Towards a standardised monitoring of ecosystem condition: A literature review on indicators and their data sources. *Ecological Indicators*, *178*, 113952. https://doi.org/10.1016/j.ecolind.2025.113952

Nyström, J., Smith, J. R., Mandle, L., Gonzalez, A., Schön, T. B., & Andermann, T. (2026). *Actionable biodiversity monitoring hinges onrepresentative data and model design*. Biodiversity. https://doi.org/10.32942/X2507T

Ovaskainen, O., Winter, S., Tikhonov, G., Abrego, N., Anslan, S., deWaard, J. R., deWaard, S. L., Fisher, B. L., Furneaux, B., Hardwick, B., Kerdraon, D., Pentinsaari, M., Raharinjanahary, D., Rajoelison, E. T., Ratnasingham, S., Somervuo, P., Sones, J. E., Zakharov, E. V., Hebert, P. D. N., … Dunson, D. (2025). Common to rare transfer learning (CORAL) enables inference and prediction for a quarter million rare Malagasy arthropods. *Nature Methods*, *22*(10), 2074–2082. https://doi.org/10.1038/s41592-025-02823-y

Peng, J., Li, J., Krause, J. R., Lyons, M. B., Murray, N. J., Schill, S. R., Roelfsema, C. M., & Asner, G. P. (2026). Global high-resolution mapping of seagrass to support conservation. *Nature*, 1–7. https://doi.org/10.1038/s41586-026-10704-3

Perino, A., Pereira, H. M., Navarro, L. M., Fernández, N., Bullock, J. M., Ceaușu, S., Cortés-Avizanda, A., van Klink, R., Kuemmerle, T., Lomba, A., Pe’er, G., Plieninger, T., Rey Benayas, J. M., Sandom, C. J., Svenning, J.-C., & Wheeler, H. C. (2019). Rewilding complex ecosystems. *Science*, *364*(6438), eaav5570. https://doi.org/10.1126/science.aav5570

Picard, J., Nungi-Pambu Dembi, M. M., Barbier, N., Cornu, G., Couteron, P., Forni, E., Gibbon, G., Lim, F., Ploton, P., Pouteau, R., Tresson, P., van Loon, T., Viennois, G., & Réjou-Méchain, M. (2024). Combining satellite and field data reveals Congo’s forest types structure, functioning and composition. *Remote Sensing in Ecology and Conservation*, *11*(2), 200–220. https://doi.org/10.1002/rse2.419

Planet, Microsoft, Natural Capital Project, & Gund Institute. (2024). *Accelerating Biodiversity and Ecosystem Reporting: Feasible and Cost-Effective Reporting With AI, Earth Observation, & Ecosystem Science*. https://www.planet.com/pulse/planet-microsoft-and-researchers-build-new-ai-focused-resource-for-environmental-reporting/

Pollock, L. J., Kitzes, J., Beery, S., Gaynor, K. M., Jarzyna, M. A., Mac Aodha, O., Meyer, B., Rolnick, D., Taylor, G. W., Tuia, D., & Berger-Wolf, T. (2025). Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nature Reviews Biodiversity*. https://doi.org/10.1038/s44358-025-00022-3

Pollock, L. J., O’Connor, L. M. J., Mokany, K., Rosauer, D. F., Talluto, M. V., & Thuiller, W. (2020). Protecting Biodiversity (in All Its Complexity): New Models and Methods. *Trends in Ecology & Evolution*, *35*(12), 1119–1128. https://doi.org/10.1016/j.tree.2020.08.015

Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., … Payne, D. (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sensing of Environment*, *239*, 111626. https://doi.org/10.1016/j.rse.2019.111626

Ren, G., Young, S. S., Wang, L., Wang, W., Long, Y., Wu, R., Li, J., Zhu, J., & Yu, D. W. (2015). Effectiveness of China's National Forest Protection Program and nature reserves: Deforestation and Protected Areas in China. *Conservation Biology*, *29*(5), 1368–1377. https://doi.org/10.1111/cobi.12561

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., & Hsiang, S. (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-24638-z

Rossberg, A. G., O'Sullivan, J. D., Malysheva, S., & Shnerb, N. M. (2024). A metric for tradable biodiversity credits quantifying impacts on global extinction risk. *Journal of Industrial Ecology*, *28*(4), 1009–1021. https://doi.org/10.1111/jiec.13518

Ryan, E. G., Drovandi, C. C., McGree, J. M., & Pettitt, A. N. (2016). A Review of Modern Computational Algorithms for Bayesian Optimal Design. *International Statistical Review*, *84*(1), 128–154. https://doi.org/10.1111/insr.12107

Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., & Hartig, F. (2021). Explainable artificial intelligence enhances the ecological interpretability of black‐box species distribution models. *Ecography*, *44*(2), 199–205. https://doi.org/10.1111/ecog.05360

Schipper, A. M., Hilbers, J. P., Meijer, J. R., Antão, L. H., Benítez‐López, A., Jonge, M. M. J., Leemans, L. H., Scheper, E., Alkemade, R., Doelman, J. C., Mylius, S., Stehfest, E., Vuuren, D. P., Zeist, W., & Huijbregts, M. A. J. (2020). Projecting terrestrial biodiversity intactness with GLOBIO 4. *Global Change Biology*, *26*(2), 760–771. https://doi.org/10.1111/gcb.14848

Schneck, J., Hawkins, F., Cox, N., Mair, L., Thieme, A., & Sexton, J. (2023). *Species Threat Abatement and Recovery in Cameroon and Kenya: Findings from a STAR assessment to support biodiversity conservation using high-resolution data*. IUCN. https://portals.iucn.org/library/node/50743

Schrodt, F., Beck, M., Estopinan, J., Bowler, D. E., Fontaine, C., Gaüzère, P., Goury, R., Grenié, M., Martins, I. S., Morueta-Holme, N., Santini, L., Hedde, M., Martin, G., Porcher, E., Si-Moussi, S., Tzivanopoulos, M., Vernham, G., Violle, C., & Thuiller, W. (2025). Advancing causal inference in ecology: Pathways for biodiversity change detection and attribution. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.70131

Seo, E., Hutchinson, R. A., Fu, X., Li, C., Hallman, T. A., Kilbride, J., & Robinson, W. D. (2021). *StatEcoNet: Statistical Ecology Neural Networks for Species Distribution Modeling* (arXiv:2102.08534). arXiv. https://doi.org/10.48550/arXiv.2102.08534

Shi, H., Norton, L., Ridding, L., Rolph, S., August, T., Wood, C. M., Qie, L., Bosilj, P., & Brown, J. M. (2026). Habitat classification from ground-level imagery using deep neural networks. *Ecological Informatics*, *95*, 103751. https://doi.org/10.1016/j.ecoinf.2026.103751

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3). https://doi.org/10.1214/10-STS330

Silvestro, D., Goria, S., Groom, B., Sterner, T., & Antonelli, A. (2025). The 30 by 30 biodiversity commitment and financial disclosure: Metrics matter. *Current Opinion in Environmental Sustainability*, *77*, 101587. https://doi.org/10.1016/j.cosust.2025.101587

Silvestro, D., Goria, S., Sterner, T., & Antonelli, A. (2022). Improving biodiversity protection through artificial intelligence. *Nature Sustainability*, *5*(5), 415–424. https://doi.org/10.1038/s41893-022-00851-6

Si-Moussi, S., Hennekens, S., Mücher, S., De Keersmaecker, W., Chytrý, M., Agrillo, E., Attorre, F., Biurrun, I., Bonari, G., Čarni, A., Ćušterevska, R., Dziuba, T., Ecker, K., Güler, B., Jandt, U., Jiménez-Alfaro, B., Lenoir, J., Svenning, J.-C., Swacha, G., & Thuiller, W. (2025). EUNIS habitat maps: Enhancing thematic and spatial resolution for Europe through machine learning. *Scientific Data*, *12*(1), 1976. https://doi.org/10.1038/s41597-025-06235-7

Sims, M. J., Stanimirova, R., Raichuk, A., Neumann, M., Richter, J., Follett, F., MacCarthy, J., Lister, K., Randle, C., Sloat, L., Esipova, E., Jupiter, J., Stanton, C., Morris, D., Melhart Slay, C., Purves, D., & Harris, N. (2025). Global drivers of forest loss at 1 km resolution. *Environmental Research Letters*, *20*(7), 074027. https://doi.org/10.1088/1748-9326/add606

Stewart, A. J., Lehmann, N., Corley, I. A., Wang, Y., Chang, Y.-C., Braham, N. A. A., Sehgal, S., Robinson, C., & Banerjee, A. (2023). SSL4EO-L: Datasets and Foundation Models for Landsat Imagery. *Advances in Neural Information Processing Systems*, *36*, 59787–59807.

Stomberg, T. T., Leonhardt, J., Weber, I., & Roscher, R. (2023). Recognizing protected and anthropogenic patterns in landscapes using interpretable machine learning and satellite imagery. *Frontiers in Artificial Intelligence*, *6*, 1278118. https://doi.org/10.3389/frai.2023.1278118

Sutherland, W. J., Burgess, N. D., Edwards, S. V., Jones, J. P. G., Soltis, P. S., Tilman, D., Allen, J. M., Andrianandrasana, H. T., Armour, C. J., August, T., Bawa, K. S., Bailey, S., Birch, T., Boersch-Supan, P. H., Cavender-Bares, J., Blaxter, M., Chaplin-Kramer, R., Daru, B. H., De Palma, A., … Robinson, G. E. (2026).

Nine changes needed to deliver a radical transformation in biodiversity measurement. *Proceedings of the National Academy of Sciences*, *123*(10), e2519345123. https://doi.org/10.1073/pnas.2519345123

Swinfield, T., Shrikanth, S., Bull, J. W., Madhavapeddy, A., & zu Ermgassen, S. O. S. E. (2024). Nature-based credit markets at a crossroads. *Nature Sustainability*, *7*, 1217–1220. https://doi.org/10.1038/s41893-024-01403-w

Tiel, N. van, Zbinden, R., Dalsasso, E., Kellenberger, B., Pellissier, L., & Tuia, D. (2025). Multi-scale and Multimodal Species Distribution Modeling. In A. Del Bue, C. Canton, J. Pont-Tuset, & T. Tommasi (Eds.), *Computer Vision – ECCV 2024 Workshops* (pp. 151–159). Springer Nature Switzerland.

Tosa, M. I., Dziedzic, E. H., Appel, C. L., Urbina, J., Massey, A., Ruprecht, J., Eriksson, C. E., Dolliver, J. E., Lesmeister, D. B., Betts, M. G., Peres, C. A., & Levi, T. (2021). The Rapid Rise of Next-Generation Natural History. *Frontiers in Ecology and Evolution*, *9*. https://www.frontiersin.org/article/10.3389/fevo.2021.698131

Vardon, M. J., & Lindenmayer, D. B. (2023). Biodiversity market doublespeak. *Science*, *382*(6670), 491–491. https://doi.org/10.1126/science.adg6823

Walker, S., Brower, A. L., Stephens, R. T. T., & Lee, W. G. (2009). Why bartering biodiversity fails. *Conservation Letters*, *2*(4), 149–157. https://doi.org/10.1111/j.1755-263X.2009.00061.x

Wallacea Trust. (2023). *Methodology for Quantifying Units of Biodiversity Gain, Version 3*. https://wallaceatrust.org/wp-content/uploads/2022/12/Biodiversity-credit-methodology-V3.pdf

West, T. A. P., Wunder, S., Sills, E. O., Börner, J., Rifai, S. W., Neidermeier, A. N., Frey, G. P., & Kontoleon, A. (2023). Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science*, *381*(6660), 873–877. https://doi.org/10.1126/science.ade3535

Westoby, M., Coleman, D., Eldridge, D. J., Kooyman, R. M., Kotowska, M., Mallen‐Cooper, M., Schrader, J., & Watson, I. W. (2026). Ecosystem valuation: Concept, criteria and terminology. *Methods in Ecology and Evolution*, *17*(5), 1480–1484. https://doi.org/10.1111/2041-210x.70279

White, T. B., Mukherjee, N., Petrovan, S. O., & Sutherland, W. J. (2023). Identifying opportunities to deliver effective and efficient outcomes from business-biodiversity action. *Environmental Science & Policy*, *140*, 221–231. https://doi.org/10.1016/j.envsci.2022.12.003

White, T. B., Viana, L. R., Campbell, G., Elverum, C., & Bennun, L. A. (2021). Using technology to improve the management of development impacts on biodiversity. *Business Strategy and the Environment*, bse.2816. https://doi.org/10.1002/bse.2816

Wunder, S., Fraccaroli, C., Bull, J. W., Dutta, T., Eyres, A., Evans, M. C., Thorsen, B. J., Jones, J. P. G., Maron, M., Muys, B., Pacheco, A., Olesen, A. S., Swinfield, T., Tegegne, Y. T., White, T. B., Zhang, H., & Zu Ermgassen,

S. O. S. E. (2025). Biodiversity Credits: An Overview of the Current State, Future Opportunities, and Potential Pitfalls. *Business Strategy and the Environment*, *34*(7), 8470–8499. https://doi.org/10.1002/bse.70018

Xu, L., Rolf, E., Beery, S., Bennett, J. R., Berger-Wolf, T., Birch, T., Bondi-Kelly, E., Brashares, J., Chapman, M., Corso, A., Davies, A., Garg, N., Gaylard, A., Heilmayr, R., Kerner, H., Klemmer, K., Kumar, V., Mackey, L., Monteleoni, C., … Tambe, M. (2023). *Reflections from the Workshop on AI-Assisted Decision Making for Conservation* (arXiv:2307.08774). arXiv. https://doi.org/10.48550/arXiv.2307.08774

Zbinden, R., Tiel, N. van, Kellenberger, B., Hughes, L., & Tuia, D. (2024). On the selection and effectiveness of pseudo-absences for species distribution modeling with deep learning. *Ecological Informatics*, *81*, 102623. https://doi.org/10.1016/j.ecoinf.2024.102623

Zhao, Q., & Hastie, T. (2021). Causal Interpretations of Black-Box Models. *Journal of Business & Economic Statistics*, *39*(1), 272–281. https://doi.org/10.1080/07350015.2019.1624293

Zhou, Y., & Ryo, M. (2025). *From Images to Insights: Explainable Biodiversity Monitoring with Plain Language Habitat Explanations* (arXiv:2506.10559). arXiv. https://doi.org/10.48550/arXiv.2506.10559

zu Ermgassen, S. O. S. E., Swinfield, T., Bull, J. W., Duffus, N. E., Macintosh, A., Maron, M., Theis, S., White, T., & Evans, M. C. (2026). Five rules for scientifically credible nature markets. *Nature Ecology & Evolution*, *10*(2), 181–192. https://doi.org/10.1038/s41559-025-02932-z

From Scalable Biodiversity Measurement to Credible Biodiversity Metrics – Supplementary Information

Authors: Douglas W. Yu[1,2]*, Fabian Ewald Fassnacht[3], Tone Birkemoe[4], Corrado Di Maria[5], Emiliya Lazarova[5], Francesca Cuomo[6], Giovanna Michelon[7], Hans Ole Ørka[4], Terje Gobakken[4], Viorel D. Popescu[8,9], Maximilian Pichler[10]

Affiliations:

[1]School of Biological Sciences, University of East Anglia, Norwich, UK
[2]Yunnan Key Laboratory of Biodiversity and Ecological Security of Gaoligong Mountain, Kunming Institute of Zoology, Kunming, China
[3]Institute of Geomatics, BOKU University, Vienna, Austria
[4]Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway
[5]School of Economics, University of East Anglia, Norwich, UK
[6]Norwich Business School, University of East Anglia, Norwich, UK
[7]Department of Economics and Management, University of Padova, Padova, Italy
[8]School of Professional Studies & Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, USA
[9]Center for Environmental Research (CCMESI), University of Bucharest, Bucharest, Romania
[10]Theoretical Ecology, University of Regensburg, Regensburg, Germany
*Corresponding author. Email: douglas.yu@uea.ac.uk

Supplementary Information 1 – Classification of the Corporate Sustainability Reporting Directive's (CSRD) Biodiversity and Ecosystem Disclosures

Background

The European Financial Reporting Advisory Group (EFRAG) is a private association that provides technical advice on the CSRD to the European Commission, most pertinently by defining the European Sustainability Reporting Standards (ESRS) that constitute the CSRD disclosures. For users, EFRAG has published an Excel spreadsheet[1] that presents the CSRD disclosures as separate datapoints, along with implementation guidance.

Methods

We downloaded the spreadsheet and combined the 125 datapoints included in ESRS Topic Standard E4 (Biodiversity and Ecosystems) with 38 relevant datapoints from ESRS 2 General Disclosure, creating a new worksheet with 163 datapoints called "ESRS 2 MDR and E4 annotations". We scored these 163 datapoints along two dimensions (*Degree of obligation to disclose* and *Data sources*) in columns F-I and grouped the 163 individual datapoints into 38 summary disclosures in column J. We then arrayed the 38 summary disclosures along the two dimensions in the "Table 0" worksheet. To improve clarity, we further grouped the 38 summary disclosures into 31 summary disclosures in the "Table 1" worksheet. Author DY carried out the first round of scoring, and authors FC and GM both independently checked the scoring and summarisation (with notes in columns L and M in the "ESRS 2 MDR and E4 annotations" worksheet.)

To show our work, each summary disclosure cell in Table 0 is linked backward (via an Excel equation) to the set of individual datapoints in "ESRS 2 MDR and E4 annotations" and linked forward (via a cell link) to one of the 31 final summary disclosures in Table 1. To follow the forward link from Table 0, click on the cell. To follow the backward link from Table 0, use the arrow keys to move the cursor into the cell and read the Excel equation. For example, the text in Table 0 cell E4 ("Resilience of SBM [strategy and business model] to biodiversity and ecosystem Risks and Opportunities") is derived via Excel equation "`='ESRS 2 MDR and E4 annotations'!J3`". Thus, navigate to column J, row 3 in the "ESRS 2 MDR and E4 annotations" worksheet. Finally, the individual datapoints in column N of "ESRS 2 MDR and E4 annotations" are linked to the full CSRD text on an EFRAG webpage.

---

[1] Finalization of Three EFRAG ESRS IG Documents (EFRAG IG 1 to 3) (2024). https://www.efrag.org/en/news-and-calendar/news/finalization-of-three-efrag-esrs-ig-documents-efrag-ig-1-to-3.

We defined four *Degrees of obligation* to disclose:

1. *Mandatory* – The company must disclose this biodiversity information.

2. *Conditionally mandatory* – The disclosure is mandatory given adoption of the relevant policy, action, or target or given disclosure of a material impact, but adoption is itself voluntary, and the procedure for impact materiality assessment is self-defined.

3. *Voluntary* – The company is free to choose whether or not to disclose.

4. *Explanations for non-adoption* – Under the CSRD's comply-or-explain model, it is mandatory to explain why any policies, actions, or targets have not been adopted. It is voluntary to explain why a timeline for adoption has not been adopted.

We defined three *Data sources*:

1. *Corporate databases and external standards* – Information available in the company's own databases or from external benchmarks suggested by the CSRD, which include financials, assets, business processes and models, activities, policies, management structure and incentives, action plans, targets, and the strategies and visions of international treaties.

2. *Static biodiversity databases* – Infrequently updated biodiversity databases, many of which are public and based in part on historical biodiversity records, such as the IUCN Red List and the World Database on Protected Areas. We define infrequent as supra-annual, since the CSRD reporting frequency is annual.

3. *Direct measures of biodiversity (remote sensing or in-field)* – Frequently updated spaceborne or airborne sensor data, such as the EU's Sentinel/Copernicus Earth Observation System, and in-field biodiversity survey data, especially novel community data collected with electronic sensors or environmental DNA(Hartig et al., 2024) and near-real-time, aggregated human observations, such as those in iNaturalist, eBird, and social media imagery. We define frequent as updated at least once a year, since the CSRD's reporting frequency is annual.

Results

Materiality assessment. – Companies are allowed to conduct a *materiality assessment* to determine which, if any, biodiversity topics need to be reported, and the CSRD does "not impose any conduct requirements in relation to [this] due diligence" (ESRS 1-4.58(CSRD, 2022)).

Mandatory disclosures requiring direct measures of biodiversity. – Following a materiality assessment process, it is mandatory to disclose a list of the “material sites in its own operations, including sites under its operational control” and break them down by their dependencies and impacts on the areas where they are located (ESRS 2 SBM-3 16 a-c).

1. These impacts include but are not limited to land degradation, desertification, soil sealing, and “operations that affect threatened species”. The ecological status of the areas must be tracked over time with reference to an “ecosystem baseline level”.

2. Any material sites near *biodiversity sensitive areas* must be reported, including any activities leading to the “deterioration of natural habitats and the habitats of species”, and “to the disturbance of the species for which a protected area has been designated” (ESRS 2 IRO-1 19 a). Biodiversity sensitive areas are defined by the CSRD(2022) as “Natura 2000 network of protected areas, UNESCO World Heritage sites and Key Biodiversity Areas (KBAs), as well as other protected areas, as referred to in Appendix D of Annex II to Commission Delegated Regulation (EU) 2021/2139”.

3. However, no metric for measuring these impacts is defined by the CSRD. Moreover, impacts (and sites) arising from value-chain partners are not included in these mandatory disclosures.

Conditionally mandatory disclosures requiring direct measures of biodiversity. – Conditionally mandatory means the disclosure is mandatory *only if the company first adopts the relevant policy, action, or target or if the undertaking discloses material impacts on biodiversity and ecosystems*. However, the CSRD does not define a standard for assessing the materiality of impacts and dependencies, adoption of the policy, action, or target to address impacts and dependencies is itself voluntary, and the metrics are left undefined, although several are suggested.

1. If policies have been adopted to track or manage impacts, risks, opportunities or dependencies on biodiversity and ecosystems, including impact drivers, the company must report metrics of its contributions to impacts, risks, opportunities, and impact drivers.

2. If policies have been adopted to address production, sourcing, or consumption from ecosystems that are managed to maintain or enhance biodiversity status, the company must regularly report biodiversity gains and losses.

3. If actions have been adopted to support the above policies, the actions and the results of those actions must be disclosed, especially ecological restoration, biodiversity offsets, and compensation to affected communities.

4. If targets have been adopted to gauge the effectiveness of policies and actions above, the company must disclose the targets, the metrics used, and performance against those targets. Example metrics suggested by the CSRD include change over time in land use, land and habitat cover, landscape connectivity, alien species distributions, ecosystem management methods (e.g. intensive vs regenerative agriculture), population or threat-status changes in threatened species, community metrics such as species compositional change and aggregated conservation statuses, and ecosystem quality.

5. Companies are allowed to use indirect information, such as peer group averages, to report on their supply chain impacts, which reduces the amount of information disclosed.

Voluntary disclosures requiring direct measures of biodiversity. – There is one fully voluntary disclosure that requires direct measures of biodiversity.

1. If policies have been adopted to address production, sourcing, or consumption specifically of raw materials, including certifications, application of the Mitigation Hierarchy, consultation, and preferential sourcing from sustainable ecosystems, the company may regularly monitor and report biodiversity status and gains and losses from supplier ecosystems.

Table S1. CSRD disclosure datapoint breakdown. Classification of the 163 CSRD disclosure datapoints in Topical Standard E4 (Biodiversity & Ecosystems) by **Degree of obligation to disclose** (rows) and **Data source** (columns), grouped into 31 summary disclosures for clarity (see Excel spreadsheet for disclosure-by-disclosure scoring). Most disclosures are conditionally mandatory or voluntary, meaning that companies are allowed to use their own materiality assessment process, which they control, to minimise disclosures. However, it is mandatory for firms to disclose their operational sites, albeit also after a materiality assessment, which provides an interface for auditing claims of biodiversity impact and state in surrounding areas.

Degree of obligation to disclose	Data source		
	Corporate data and external standards	**Static biodiversity databases**	**Direct measures of biodiversity and ecosystems (remote sensing or in-field)**
Mandatory disclosures	Resilience of SBM to biodiversity and ecosystem Risks and Opportunities, and quantitative and qualitative information on how biodiversity and ecosystem Risks, Opportunities, and Dependencies will affect the undertaking's finances		
	Description of processes to assess materiality of biodiversity and ecosystem-related IRODs, including consultations with affected communities and mitigation measures that are required by EU legislation or IFC Performance Standard 6		
	Activities related to own sites that lead to deterioriation of "natural habitats and the habitats of species", negatively affect biodiversity-sensitive areas, or disturb the species for which a protected area was designated		
	List of all material sites owned or under operational control ('own sites')	Own sites that are in or near biodiversity-sensitive areas →	Breakdown of own sites by their Dependencies and Impacts on the areas where they are located (including land degradation, desertification, soil sealing, and threatened species), and the ecological statuses of those areas, with reference to baseline levels. Identify any biodiversity-sensitive areas impacted.
Conditionally mandatory disclosures (Mandatory to disclose if policy, action, target, metric, or method has been adopted or if an impact is disclosed, but the adoption is voluntary, and the impact assessment methods are self-defined)	Policies adopted to track and manage IRODs related to biodiversity and ecosystems, including contributions to impact drivers on biodiversity loss, social consequences, and inputs and outputs along the value chain. Disclosure of any third-party standards and availability of policies to stakeholders	→	Metrics of the undertaking's contributions to IROs and to impact drivers of land-use change, freshwater-use change, and/or sea-use change. Description of methodologies and assumptions behind metrics.
	Policies adopted to address production, sourcing, or consumption from ecosystems that are managed to maintain or enhance biodiversity status	→	Regular monitoring and reporting of biodiversity status and gains or losses
	Policies adopted for the protection of biodiversity-sensitive areas near undertaking's operational sites, for sustainable agriculture, for sustainable oceans practices, and for addressing deforestation		
	Key actions taken and/or planned to support the adopted policies in relation to biodiversity IRODs, including scope, and time horizon. Any use of biodiversity offsets, including key performance indicators, costs, and use of local ecological knowledge	→	Quantitative and qualitative information on progress and results of key actions taken, including ecological restoration, biodiversity offsets, and/or compensation to affected communities
	Targets used to gauge effectiveness of policies and actions. Methodology used to set targets: scope, baseline value and/or year, time period, milestones or interim targets, methods and scientific evidence base for defining targets, stakeholder involvement in target, changes in targets, reference to ecological thresholds, global or national policies or laws, the undertaking's material IROs, and the Mitigation Hierarchy	→	Description of performance against target, including how target is monitored and reviewed, metrics used, whether progress is in line with initial plans, and analysis of trends or significant changes in progress towards achieving the target.
	Financial resources allocated to actions and action plans. Do company actions involve significant Capex or Opex, and how do they relate to financial statements? How do company actions involve stakeholders? Will the actions lead to negative Impacts on stakeholders and/or on biodiversity and ecosystems? Are actions one-time initiatives or a systematic practice? Are actions part of a collective action?		
	Number and areas of owned, leased, or managed sites that have been identified in the ESRS2.IRO-1 process as having negative Impacts on protected areas, biodiversity-sensitive areas, or key biodiversity areas		
	Analysis of land use based on Life Cycle Assessment		
Voluntary disclosures	Transition plan to align SBM with local, national, and global biodiversity and ecosystem goals		
	Plans to minimise or mitigate unavoidable negative Impacts on priority ecosystem services		
	Use of scenario analysis to identify and assess material Risks and Opportunities		
	Policies adopted to address production, sourcing, or consumption specifically of raw materials, including third-party certifications, application of the Mitigation Hierarchy, consultation, and preferential sourcing from sustainable ecosystems	→	Regular monitoring and reporting of biodiversity status and gains or losses in the sustainable ecosystems
	Has the Mitigation Hierarchy been applied, and if so, how?		
	Do targets aid transition to sustainable activity ("Substantial Contribution" criteria)?		
	Metrics chosen to measure impact drivers and IROs. It is mandatory to report metrics (and their methodologies and assurance providers) if the undertaking concludes that it has contributed to impact drivers or IROs, but the metrics are not defined Example metrics suggested by the CSRD include land-use, land-cover change, landscape connectivity, alien species distributions, ecosystem management methods (e.g. regenerative agriculture), population or threat-status changes in threatened species, community metrics such as species compositional change, and ecosystem quality.		
	Products and services at risk from biodiversity and ecosystem change. Methods used to estimate financial amounts.		
Mandatory explanations for non-adoption	Explanations for not yet having adopted policies or actions		
	In the absence of measurable, outcome-oriented targets, are alternative methods being used to track the effectiveness of policies or actions?		
Voluntary explanations for non-adoption	Explanations for not yet having adopted measurable, outcome-oriented targets		
	Timeframes for adopting policies, actions, and/or targets		

Glossary:
Undertaking: The company making the disclosures
SBM: Strategy and Business Model
IROD: Impacts, Risks, Opportunities, & Dependencies
IFC: International Finance Corporation

Supplementary Information 2 – The CSRD’s biodiversity disclosures will have low credibility with current auditing capacity

The aim of this section is to introduce a principal-agent model highlighting a firm’s incentives to measure and truthfully report its biodiversity impact. We start from a model of voluntary disclosure, i.e. a situation in which firms might decide to reveal (some of) their impacts on biodiversity to accrue economic benefit.

A simple model of voluntary disclosures. – Biodiversity is a complex and multidimensional concept, as are the impacts of the firm’s operations on it. Here, we start by assuming that the “footprint” (= negative biodiversity impact), θ, of a firm is drawn from a distribution Θ over the closed interval $[\underline{\theta},\ \overline{\theta}]$. The smaller the θ, the lower the footprint of the firm on biodiversity. We also allow for the possibility that the firm’s activity might have positive side effects on biodiversity and therefore, we let $\underline{\theta} < 0$.

The firm cannot readily observe θ, so to learn about its actual footprint it needs to invest resources to measure the biodiversity impact. The firm has at its disposal an estimator, $\widehat{\theta}_f$, which is unbiased, has variance σ_f^2, and can be used at a fixed cost, k. Depending on whether the firm decides to measure its impact, its information set becomes: $I \in \{\{\varnothing\}, \{\ \widehat{t}_f\}\}$ where $\widehat{t}_f$ is the estimate it obtains by using the metric $\widehat{\theta}_f$.

In the second step, the firm decides whether to issue a report on its footprint. Denoting the firm’s disclosures by $\widehat{t}_R$, we let the firm choose whether to issue a report, and if so, its degree of truthfulness: $\widehat{t}_R \in [\underline{\theta},\ \overline{\theta}] \cup \{\varnothing\}$, given its information set I.

The firm potentially benefits from disclosing its footprint in two ways. Firstly, low-footprint firms may benefit from the presence of sustainability-sensitive customers, who are willing to pay more for greener products. We assume that the firm has market power in its product market and that the premium it can charge over marginal cost decreases with its (reported) biodiversity impact, i.e. if no report is issued, consumers assume the firm’s impact is $\overline{\theta}$. Formally, we capture this aspect by assuming that the firm’s revenues are a decreasing function of the reported footprint, i.e. $R(\widehat{t}_R)$, such that $R^{'} < 0$ and $R^{''} > 0$.

$$\widetilde{R}\left(\widehat{t}_R\right)=\begin{cases}R\left(\widehat{t}_R\right), & \text{if } \widehat{t}_R\in[\underline{\theta},\ \overline{\theta}]\\ R\left(\overline{\theta}\right), & \text{if } \widehat{t}_R=\{\varnothing\}\end{cases}$$

We assume that with some probability, $\alpha_c \in (0,1)$, a consumer watchdog, whose objective is to protect consumer interests by monitoring and reporting on firms' activities, can utilise an alternative estimator $\widehat{\theta}_C$ to conduct independent measurements on the firm's reported biodiversity impact, should the firm issue a report $\widehat{t}_R \in \left[\underline{\theta}.\ \overline{\theta}\right]$. We assume that, like the estimator available to the firm, the watchdog's estimator is unbiased and has variance $\sigma_C^2 \geq \sigma_f^2$. We assume that, given a firm's report $\widehat{t}_R \in \left[\underline{\theta}.\ \overline{\theta}\right]$ and the watchdog estimate $\widehat{t}_C$, if the probability that the firm has under-estimated its impact exceeds the threshold $\overline{\alpha}$, with $\overline{\alpha}=\mathrm{Prob}(\theta \geq \widehat{t}_R|\widehat{t}_C)$, the watchdog launches a whistleblowing campaign, causing the firm to incur reputational damage equal to F_C. Thus, from the firm's perspective, the consumer watchdog's whistleblowing decision is summarised by

$$\Psi\left(\widehat{t}_C,\ \widehat{t}_R\right)=\begin{cases}1, & \text{if } \Pr\left(\theta>\widehat{t}_R|\widehat{t}_C\right)>\overline{\alpha}\\ 0, & \text{if } \Pr\left(\theta>\widehat{t}_R|\widehat{t}_C\right)\leq\overline{\alpha} \text{ or } \widehat{t}_R=\varnothing\end{cases}$$

From the perspective of the firm, knowing the monitoring technology that the watchdog has at its disposal, the probability that the watchdog will whistleblow, if it conducts a measurement, can be summarised by the following expression if the firm has decided not to conduct its own measurement and issue a report $\widehat{t}_R$:

$$\Omega_0\left(\widehat{t}_R\right)=\mathrm{Prob}\left(\Psi\left(\widehat{t}_C,\ \widehat{t}_R\right)=1\right)\ ;$$

and instead it is summarised by the following conditional probability if the firm has conducted its own biodiversity impact measurement:

$$\Omega_1\left(\widehat{t}_R,\widehat{t}_f\ \right)=\mathrm{Prob}\left(\Psi\left(\widehat{t}_C,\ \widehat{t}_R\right)=1|\widehat{t}_f\right).$$

Secondly, the firm may benefit from accessing a competitive market for ethical finance. To do so, however, it needs to prove its sustainability credentials. In particular, we assume that it needs to show that its (reported) footprint, $\widehat{t}_R$ is below a given benchmark $\widetilde{\theta}$, with $\widetilde{\theta}\in\left(\underline{\theta},\ \overline{\theta}\right)$. Access to ethical finance implies a lower financing cost for the firm, and the cost of financing decreases with the footprint. In this respect, the cost function of the firm decreases with the (positive) difference $(\widetilde{\theta} - \widehat{t}_R)$ and equals a higher constant otherwise, i.e.:

$$\widetilde{C}\left(\widetilde{\theta} - \widehat{t}_R\right) = \begin{cases} C(\widetilde{\theta} - \widehat{t}_R), & \text{if } \widetilde{\theta} - \widehat{t}_R \geq 0 \\ \overline{C} > C(0), & \text{if } \widetilde{\theta} - \widehat{t}_R < 0 \text{ or } \widehat{t}_R = \{\varnothing\} \end{cases}$$

where we assume, intuitively, that $C^{'} < 0$ and $C^{''} > 0$. When the firm accesses ethical finance, its reporting may be subject to auditing by funders, to reduce the incentives to misreport its biodiversity impacts. We let the probability that the firm's report is audited be $\alpha_f \in (0, 1)$. We will assume that by law the firm needs to make all its paperwork accessible to financial auditors. The financial audit therefore consists of verifying that the firm's reported biodiversity impact is truthful, i.e. financial auditors check that $\widehat{t}_R = \widehat{t}_f$. If the audit establishes that the firm has under-reported its impact (Note that in this framework the firm has no incentive to over-report its impact), then the firm needs to pay a large monetary penalty of size $F_f > F_C$.

We note that the two external auditing channels play quite different roles. The watchdog forms a probabilistic inference from its own noisy estimate and whistleblows when the gap between its estimate and the firm's report is large enough to make misreporting sufficiently likely. By contrast, the ethical-finance investor audits the truthfulness of reporting relative to the firm's own measurement process: if the firm reports without having measured, that is misreporting; if the firm has measured but reports anything other than its own measurement, that is also misreporting.

Given the discussion above, it follows that the firm decides whether to perform any measurement and chooses its disclosures to maximise its profits, keeping in mind the possibility of both an independent measurement and whistleblowing threat by the consumer watchdog and of an audit with hard evidence verification by the ethical financial investor, its own measurement cost, and the costs of being found to have misrepresented or even fabricated its footprint measurement. We therefore write the firm's maximisation problem over two choice variables $m \in \{0, 1\}$ which is whether it should conduct own measurement ($m = 1$) or not ($m = 0$) and what report to issue: as follows, $\widehat{t}_R \in [\underline{\theta}.\ \overline{\theta}] \cup \{\varnothing\}$ based on the following payoff function:

$$\max\ _{m \in \{0,1\},} \left\{ \begin{array}{l} (1-m) \max_{\widehat{t}_R \in [\underline{\theta}.\ \overline{\theta}] \cup \{\varnothing\}} \left[\widetilde{R}\left(\widehat{t}_R\right) - \widetilde{C}\left(\widehat{t}_R\right) - \alpha_C F_C \Omega_0\left(\widehat{t}_R\right) - \alpha_F F_F \mathbf{I}\left(\widehat{t}_R < \widetilde{\theta}\right) \mathbf{I}\left(\widehat{t}_R \neq \varnothing\right)\right] + \\ m[E_{\widehat{t}_f}\left(\max_{\widehat{t}_R \in [\underline{\theta}.\ \overline{\theta}] \cup \{\varnothing\}} \left\{\widetilde{R}\left(\widehat{t}_R\right) - \widetilde{C}\left(\widehat{t}_R\right) - \alpha_C F_C \Omega_1\left(\widehat{t}_R,\ \widehat{t}_f\right) - \alpha_F F_F \mathbf{I}\left(\widehat{t}_R < \widetilde{\theta}\right) \mathbf{I}\left(\widehat{t}_R \neq \widehat{t}_f\right)\right\}\right) - k] \end{array} \right\}$$

where the $\mathbf{I}(\cdot)$'s denote indicator functions that equal 1 when the conditions in the brackets are satisfied, and 0 otherwise; and $E_{\widehat{t_f}}$ denotes an expectation conditional on the observed estimate of the firm's biodiversity impact.

The way to analyse this model is by backward induction. Because we have kept the revenue function, the financing-cost function, and the distributions of the signals in general form, the model does not admit a fully explicit closed-form solution; instead, we focus our discussion on cutoff rules.

Measurement and truthful disclosure

We focus our discussion on the case in which the firm chooses to measure its biodiversity impact at cost k, and whether to disclose it truthfully.

Let us fix a realised internal estimate t_f. For any under-reporting of the biodiversity impact $t_r < t_f$, we can define the firm's gain from under-reporting as

$$G(t_r, t_f) = [R(t_r) - R(t_f)] + [\widetilde{C}(t_f) - \widetilde{C}(t_r)],$$

where $\widetilde{C}(x)$ denotes the financing cost associated with a report x: it equals $C(\widetilde{\theta} - x)$ if $x \leq \widetilde{\theta}$, and $\bar{C}$ otherwise**.** Let us denote the extra expected sanction from under-reporting as

$$S(t_r, t_f) = \alpha_C F_C[\Omega^1(t_r, t_f) - \Omega^1(t_f, t_f)] + \alpha_F F_F \ \mathrm{I}\{t_r \leq \widetilde{\theta}\}.$$

Hence, truth-telling t_f is optimal if, for every feasible under-reporting $t_r < t_f$,

$G(t_r, t_f) \leq S(t_r, t_f).$ Then no downward misreport is profitable, so truthful disclosure weakly dominates every active misreport after measurement. Since the firm also has no incentive to over-report its footprint, the reporting problem after measurement reduces to a comparison between truthful disclosure and non-disclosure.

Let

$$\Delta(t_f) = [R(t_f) - R(\bar{\theta})] + [\bar{C} - \widetilde{C}(t_f)] - \alpha_C F_C \ \Omega^1(t_f, t_f),$$

be the payoff advantage of truthful disclosure over non-disclosure. If $\Delta(t_f)$ is continuous and strictly decreasing in t_f, then there exists a unique cutoff t^* such that a measuring firm truthfully discloses if and only if $t_f \leq t^*$, while it chooses non-disclosure when $t_f > t^*$. The cutoff is characterised by the indifference condition $\Delta(t^*) = 0$. Finally, the firm chooses to measure if and only if the expected value of entering this stage-two problem, net of the fixed cost k, is at least as large as the best payoff attainable without measurement.

Sketch of proof. We need to proceed by backward induction. We start with the reporting stage and fix t_f. In our model, any profitable active misreport must be a downward misreport, because revenues are higher when reported biodiversity harm is lower, and access to ethical finance is easier when the reported harm is below $\widetilde{\theta}$. The condition $G(t_r,\ t_f) \leq S(t_r, t_f)$ states exactly that, for every possible under-report t_r, the gain from misstating the footprint cannot be larger than the increase in expected sanctions. Hence no downward deviation can improve on truthful disclosure. Since upward deviations are never beneficial, truthful disclosure is the best active report available after measurement.

Once active misreporting has been ruled out, the firm compares truthful disclosure with silence. The payoff difference is $\Delta(t_f)$. If $\Delta(t_f)$is continuous and strictly decreasing, it can cross zero at most once, which implies a unique cutoff t^*. Firms with sufficiently low realised biodiversity impact disclose truthfully; firms with sufficiently high realised biodiversity impact choose to remain silent. The measurement decision in stage one then compares the expected value of this optimal stage-two behaviour with the fixed cost k and the best payoff obtainable without measurement. ■

It is instructive to think about how the likelihood of measurement and truthful reporting changes with the parameters of the formal environment. A useful way to formalise the comparative statics is to distinguish between two margins of adjustment in the model: first, the measurement margin, that is, whether the firm chooses to incur the fixed cost of learning its biodiversity impact; and second, the truth-telling margin, that is, whether a firm that has measured prefers to disclose truthfully rather than understate or remain silent. This distinction matters because some parameters affect only the *ex-ante* value of becoming informed, while others primarily affect the profitability of misreporting once information has been acquired. In the present framework, measurement becomes more attractive when the option value of private information rises relative to its cost, whereas truthful disclosure becomes more attractive when the expected sanctions associated with understatement rise relative to the private gain from appearing greener or from accessing ethical finance on more favourable terms.

Corollary (comparative statics). Holding fixed the revenue function, the financing-cost function, the support of biodiversity impacts, and the structure of watchdog and audit monitoring, the likelihood that the firm conducts a measurement is strictly decreasing in the fixed measurement cost k. A higher k reduces one-for-one the *ex-ante* net payoff from becoming informed and therefore makes it less likely that the firm will measure. By contrast, k does not directly alter the profitability of truthful disclosure relative to misreporting conditional on measurement; it affects the reporting stage only indirectly by changing whether the firm reaches that stage at all.

Next, an increase in the expected reputational sanction from watchdog whistleblowing, captured by $\alpha_C F_C$, tends to strengthen the truth-telling margin by making understatement more costly in expectation. The larger is $\alpha_C F_C$, the smaller is the net gain from reporting an artificially low biodiversity footprint. Its effect on the measurement margin is generally positive when the main benefit of measurement is that it allows the firm to decide more carefully whether credible disclosure is worthwhile. In that case, stronger reputational discipline raises the value of being informed before the firm commits to a disclosure strategy. However, the effect need not be mechanically monotone for all types, because very severe watchdog sanctions may also make disclosure unattractive for firms with sufficiently poor realised impacts.

Next, an increase in the expected audit sanction from the ethical-finance investor, captured by $\alpha_F F_F$, makes truthful disclosure more likely by directly reducing the profitability of misreporting whenever the firm seeks to qualify for ethical finance. The stronger this sanction, the less attractive it becomes to report a value below the true internal estimate to cross the ethical-finance threshold. This also raises the value of measurement itself, because informed and internally consistent reporting becomes more valuable in a regime where access to ethical finance is policed more strictly. Thus, unlike the reputational channel, a rise in $\alpha_F F_F$ tends to strengthen both the incentive to measure and the incentive to report truthfully.

Finally, a larger precision advantage of the firm's own estimator over the watchdog's estimator — equivalently, a lower σ_f^2relative to σ_C^2 — increases the attractiveness of measurement, because the firm's private information then becomes more informative than the signal available to outside monitors. This raises the option value of internal measurement. Its effect on truthful disclosure is more subtle. On the one hand, a more precise internal signal improves the value of informed and accurate reporting. On the other hand, if the watchdog's signal becomes relatively noisier, external detection of understatement becomes weaker, and the deterrent effect of reputational sanctions may be softened. Conversely, if the watchdog becomes relatively more precise, the expected sanction from misreporting strengthens, which makes truth-telling more attractive, but the firm's private informational advantage from measuring becomes smaller. The net effect of relative precision on truthful disclosure therefore depends on the balance between the information value of internal measurement and the disciplinary role of external scrutiny.

Sketch of proof. The proof again proceeds by backward induction. Start with the reporting stage conditional on a realised internal estimate t_f. The profitability of any downward misreport depends on the comparison between its private gain — through higher revenues and, where relevant, cheaper access to ethical finance — and its expected sanction — through a higher probability of whistleblowing and, if the report is used to obtain ethical finance, a possible audit penalty when the report does not match the firm's own internal record. A higher $\alpha_C F_C$ or a higher $\alpha_F F_F$ raises the expected cost of understatement and therefore weakens the incentive to misreport, making truthful disclosure relatively more attractive. This establishes the comparative statics for the truth-telling margin.

Turn next to the measurement stage. The *ex-ante* value of measurement is the expected value of entering the reporting problem with private information, net of the fixed cost k. Since k enters linearly and negatively, a higher k unambiguously lowers the likelihood of measurement. A higher expected audit sanction raises the value of having an internal record that can support a credible report and therefore increases the return to measurement. A higher expected reputational sanction can also raise the return to measurement by making uninformed or opportunistic disclosure riskier, although this effect is weaker when harsher scrutiny discourages disclosure altogether for firms with poor realised impacts. Finally, a fall in σ_f^2 relative to σ_C^2 increases the value of measurement by improving the quality of the firm's private signal relative to that of the watchdog. The comparative statics therefore follow from how each parameter shifts either the *ex-ante* value of information or the ex-post profitability of under-reporting.

What happens with mandatory disclosures? – How mandatory disclosures like in the CSRD change the model discussed above very much depends on the type of mandate we have in mind. In the first regime, measurement remains voluntary, but disclosure is mandatory. In the second regime, both measurement and disclosure are mandatory.

The economic logic is different because, in the first case, the firm can still choose whether to acquire internal information before it is forced to report, whereas in the second case, the firm is compelled both to know and to disclose its biodiversity impact. In both cases, however, the key issue is whether enforcement is strong enough to make mandatory reporting credible rather than merely formal.

Corollary 1 (mandatory disclosure with voluntary measurement). Suppose disclosure of biodiversity impact is mandatory, but the firm is not required to conduct its own measurement. The firm must therefore issue a report, but it still chooses whether to incur the fixed cost k to obtain its own internal estimate t_f. In this regime, the firm's stage-two problem no longer includes non-disclosure as an option; instead, if it has measured, it compares truthful disclosure with the best feasible downward misreport, and if it has not measured, it chooses the most profitable report available without private information. Truthful reporting after measurement is optimal if and only if, for every realised internal estimate t_f and every feasible downward deviation $t_r < t_f$, the private gain from understating biodiversity harm is weakly smaller than the increase in expected sanctions from watchdog whistleblowing and, whenever the report is used to obtain ethical finance, from the expected audit penalty associated with reporting a value that does not match the firm's internal record. Under this regime, mandatory disclosure weakly raises the amount of reporting relative to the voluntary-disclosure benchmark, but it does not necessarily improve reporting accuracy. If expected sanctions are weak, firms that would previously have remained silent may now comply formally while understating their biodiversity impact. The effect on measurement is ambiguous. On the one hand, mandatory disclosure can increase the value of measurement because private information helps the firm decide how to report when silence is no longer available. On the other hand, if the firm can submit a report without internal measurement and enforcement against unsupported reporting is weak, measurement may become less attractive because remaining uninformed preserves flexibility and avoids creating an internal record against which a future report could be checked. Thus, in this regime, mandatory disclosure increases reporting but improves information quality only when sanctions are sufficiently strong to deter strategic understatement and when internal information remains valuable for disclosure decisions.

Sketch of proof. Relative to the voluntary regime in the base case, the only strategic option removed is non-disclosure. The firm must therefore report something regardless of type. If the firm measures, the relevant comparison is between truthful reporting and under-reporting; if it does not measure, it chooses the best uninformed report available. The incentive-compatibility condition for truthful reporting is therefore the same as in the earlier proposition, except that silence is no longer part of the choice set. This tends to strengthen the temptation of poor types to understate, because they can no longer avoid the market and financing consequences of a bad signal by withholding disclosure. Hence mandatory disclosure alone does not guarantee credibility. The measurement decision becomes a comparison between the fixed cost k and the value of private information when reporting is unavoidable. Because private information can either improve reporting decisions or constrain the firm by exposing inconsistency, the net effect on measurement is not necessarily monotone. ■

Corollary 2 (mandatory disclosure with mandatory measurement). Suppose now that both disclosure and internal measurement are mandatory. The firm must incur the measurement cost k, obtain its own internal estimate t_f, and issue a report. In this regime, the measurement margin disappears: every firm becomes informed before reporting. The only remaining strategic issue is whether the report coincides with the firm's measured estimate or understates it. Truthful reporting is therefore optimal if and only if, for every realised internal estimate t_f and every feasible downward deviation $t_r < t_f$, the private gain from under-reporting is weakly smaller than the increase in expected sanctions from watchdog whistleblowing and, where relevant, the expected audit penalty associated with reporting a value that does not match the firm's internal record. Relative to the regime with mandatory disclosure but voluntary measurement, this stronger regime generally improves the informational content of reports because all firms are forced to acquire internal information before reporting. It also makes audit-based enforcement more effective, because every firm has an internal record against which its disclosure can potentially be checked. However, mandatory measurement does not by itself ensure truthful reporting. If expected sanctions remain weak, firms may still understate their biodiversity impact even though they are now informed. The central implication is that mandatory measurement removes the possibility of strategic ignorance, while mandatory disclosure removes the possibility of strategic silence; what remains is strategic misreporting, and the extent of that distortion depends entirely on the strength of monitoring and enforcement.

Sketch of proof. Once both measurement and disclosure are mandatory, the firm's problem is reduced to a one-stage reporting problem conditional on a realised internal estimate t_f. Since the firm must measure, it always has a private signal and always creates an internal record. Since it must disclose, silence is no longer feasible. The only deviation from truthful reporting is therefore downward misreporting. The profitability of such a deviation is determined by the same comparison as before: the benefit of appearing greener and, if relevant, qualifying more easily for ethical finance, versus the increased expected cost of whistleblowing and the increased expected audit loss from reporting something different from the firm's own records. If the sanction condition holds, truthful reporting follows; if it does not, mandatory measurement and mandatory disclosure together simply produce mandatory but potentially distorted reporting. The difference relative to the previous corollary is that firms can no longer avoid the cost of becoming informed, so strategic ignorance disappears even though strategic understatement may remain. ■

The intuition can be summarised quite compactly. With mandatory disclosure but voluntary measurement, regulation removes silence but not ignorance, so firms may respond either by learning more or by reporting opportunistically without measuring, depending on the enforcement environment. With mandatory disclosure and mandatory measurement, regulation removes both silence and ignorance, so the only remaining distortion is misreporting. This second regime therefore produces more informative reporting on average, but only strong sanctions can ensure that the information disclosed is also truthful. In other words, mandatory disclosure increases the quantity of information released, mandatory measurement increases the informational basis for reporting, and only effective monitoring determines whether the resulting reports are credible.

What if firms can also adjust their activities? – This model has omitted several extensions, including the possibility of firms modifying their operations in response to disclosure requirements. If disclosures are mandatory but auditing is *lax*, it might pay for firms to increase their impact, while reporting a smaller footprint. On the other hand, if disclosures are mandatory and auditing *effective*, which is the best-case scenario for the CSRD, firms would have an incentive to invest in reducing their footprint.

Four criteria for credible biodiversity metrics. – *Low marginal cost* enters as the audit probabilities α_F and α_C, which would rise as independent verification becomes cheaper. *Low dispersion* enters through the relative precision of the firm's and watchdog's estimators (σ_f^2, σ_c^2). *Sufficient information content* is implicit in the assumptions that $R' < 0$ and $C' < 0$: consumers and investors only condition behaviour on the reported metric if it carries genuine ecological signal. *Parsimony* is not formally modelled. The model assumes a single metric θ for tractability, but a richer model allowing the firm to select from a menu of metrics would enable cherry-picking, as we argue in the main text, undermining any credibility gains achieved by the other three criteria.

Supplementary Information 3 – Deep Learning Species Distribution Models

Here we summarise how DL-SDMs are making progress toward scalable biodiversity measurement and highlight capability gaps that require further methodological development (Figure 1), and in Table S2, we describe twenty recent studies that primarily use remote-sensing environmental covariates to predict species distributions or summary metrics.

Environmental sensing. – For most studies, remotely sensed environmental predictors are drawn from satellite platforms. Sentinel-2 and Landsat provide multispectral imagery at 10–30 m resolution, sometimes combined into time-series composites to capture vegetation phenology(Estopinan et al., 2022; Leblanc et al., 2025). Other studies use airborne imagery at 0.2–1 m resolution(Deneu et al., 2022; Gillespie et al., 2024; Hu et al., 2025; Li et al., 2024). Global bioclimatic, soil, and topographic rasters serve as complementary predictors, and multimodal models that combine these tabular covariates with imagery have been found to outperform unimodal (tabular-only) models(Dollinger et al., 2024; Tiel et al., 2025).

Species sensing. – The most common sources of species training data are records from open databases, which include presence-only records from GBIF, iNaturalist, and GeoLifeCLEF(Cole et al., 2020; Deneu et al., 2021; Gillespie et al., 2024), presence-absence checklists from eBird and eButterfly(Davis et al., 2023; Prudic et al., 2017; Teng et al., 2023), and plot-based vegetation surveys such as the European Vegetation Archive and GeoLifeCLEF(Leblanc et al., 2025). These provide massive sample sizes but are dominated by presence-only data, geographically biased toward Western Europe and North America, and taxonomically concentrated on plants and birds.

Less commonly, novel community data have been used in the form of metabarcoded arthropods(Baggström et al., 2025; Li et al., 2024) and camera trap detections(Haucke et al., 2025). Although numbers of species records are much lower than in the open databases, these methods generate up-to-date species records from taxa and locations that are underrepresented in the open databases.

Modelling. – To date, the most common approach uses convolutional neural networks (CNNs) to analyse satellite imagery and recognise spatial features ranging from land cover to complex fragmentation patterns surrounding a species' location. Because this method captures landscape context as well as local environmental conditions, it improves prediction performance relative to approaches that only measure the environment at single points(Deneu et al., 2021, 2022). The visual information is sometimes combined with tabular data like soil and climate layers. The tabular data are handled by a different deep learning method called a multilayer perceptron (MLP), and the CNN and MLP outputs are combined(Davis et al., 2023; Gillespie et al., 2024; Hu et al., 2025).

The combined output summarises the complex environmental information at and around each location in a long vector of numbers, known as an *embedding*. This embedding provides the direct variables that the model's final classification layers use to calculate the probability of a species being present.

During training, the model repeatedly compares its calculated presence probabilities to observed species presences and absences (using loss functions) and adjusts the parameters (model weights) of the CNN, MLP, and the classification layers to minimise error. Through this iterative process, the model learns to recognise the visual and spectral characteristics of the species' preferred environment. By training on hundreds to thousands of species at once, the model can also learn general environment–occurrence relationships, which further improves its ability to map biodiversity(Deneu et al., 2021; Estopinan et al., 2022).

Outputs. – The proximate output is a stack of species distribution maps, each predicting the likelihood of a species occurring at each grid cell. Published maps have ranged in resolution from 10 m(Leblanc et al., 2025) to 2.9 km grid cells(Davis et al., 2023). One of the largest DL-SDMs to date has mapped 5,558 plant species across 5.5 billion grid cells in Europe(Leblanc et al., 2025). The individual maps can be summarised into maps of species richness and composition(Hu et al., 2025), and Leblanc et al.(2025) processed their DL-SDM outputs with a transformer model to classify sites to EUNIS habitat classes. There are also studies that, instead of modelling individual species, have directly trained on and predicted biodiversity summary metrics like species richness(Baggström et al., 2025), forest types(Picard et al., 2024), seagrass extent(Giménez-Romero et al., 2024; Peng et al., 2026), and protected-versus-anthropogenic land cover(Stomberg et al., 2023).

DL-SDM extensions that improve performance or efficiency

Studies have extended DL-SDMs in various ways to improve predictive performance and training efficiency.

Accounting for false negative error. – Species surveys are notorious for producing false negatives, sites where a species is truly present but not detected. In classical ecological statistics, false-negative error is accounted for with occupancy models(Mackenzie et al., 2002; Tyre et al., 2003), which use separate submodels to estimate species *presence* versus species *detectability* given presence. Seo et al.(2021) similarly embedded a detection submodel in their DL-SDM, and Davis et al.(2023) propose that deep learning could learn to account for false-negative error by including covariates that might be affecting species detection, such as survey effort.

Class-imbalance-aware loss functions. – Most species are rare, so species training data are severely class-imbalanced in the sense that absences outweigh presences, and a small number of common species dominates the training signal. Using loss functions that account for this, such as by downweighting absences in low-sampling-effort sites, has been shown to improve predictions(Gillespie et al., 2024; Kellenberger et al., 2026; Zbinden et al., 2024).

Proximal sensing of local environmental covariates. – Low-cost proximal sensors, such as passive acoustic recorders, cameras, and terrestrial or airborne lidar can improve predictive performance by providing high-resolution environmental measurements at individual sites (Figure 1). For instance, Haucke et al.(2025) found that adding a single animal-free camera image from each of the thousands of camera-trap sites across the US improved predictive performance over models that used only satellite imagery. Similarly, Li et al.(2024) and Baggström et al.(2025) included fine-grained structural vegetation covariates calculated from airborne lidar in their models predicting arthropod species distributions across a working forest in Oregon and arthropod species richness across Sweden, respectively. The advantage of proximal sensing is the added information from microhabitat features that are not visible to satellites, but the disadvantage is that the proximal measurements must be taken anywhere that one wants to make a prediction, which requires site visits, thereby reducing scalability (and raising marginal costs) relative to models relying solely on remote-sensing predictors.

Conditional prediction based on species correlations. – Another way to improve model performance when predicting species at new sites is to exploit the species correlations learned from the biodiversity training data. These correlations carry information about unmeasured environmental features, biotic interactions, and assembly rules(Poggiato et al., 2021). Independent observations of some species at new sites, via proximal sensing or separate biodiversity surveys, can be used to condition predictions of their correlated species at those sites(Abdelwahed et al., 2024, 2026). However, as with proximal sensing, scalability is reduced.

Both proximal sensing and conditional prediction thus trade low marginal cost (criterion 1) for higher information content (criterion 3) and lower dispersion (criterion 2), and are therefore more justifiable when the decision stakes are high, such as at restoration sites or where a legally protected species might be present.

Transfer learning. – Transfer learning from large ecological or remote-sensing datasets can help overcome limited biodiversity training data. For example, Bourel et al.(2025) pre-trained a CNN on a larger *presence-only* fish dataset before fine-tuning it on a smaller fish *abundance* dataset, improving abundance prediction over a non-pre-trained DL-SDM. More broadly, when satellite or aerial imagery is analysed by CNN models, early training rounds are devoted to parsing images into visual elements. Pre-training on large generic image datasets such as ImageNet fast-forwards this process, allowing models to learn ecologically relevant features with less biodiversity training data(Kellenberger et al., 2026). Geospatial foundation models trained on massive satellite imagery archives extend this principle, providing ready-to-use embeddings for DL-SDMs(Brown et al., 2025; Rolf et al., 2021; Stewart et al., 2023) and potentially reducing training data requirements generally, particularly for new regions and taxa.

Capability Gaps

Important capabilities remain underdeveloped: information sharing, causal attribution, and extrapolation to novel conditions.

Information sharing. – Species share phylogenetic, trait, and spatial similarities, and models that share such information during model training can improve predictions. DL-SDMs already implicitly share information when trained on multiple species at once, improving predictions(Davis et al., 2023; Harrell et al., 2025), including for data-deficient species(Dinnage, 2024). Explicitly incorporating phylogenetic or trait structure into DL-SDMs has been proposed(Harrell et al., 2025) but not, to our knowledge, implemented, though a Bayesian joint species distribution model (JSDM) has used phylogenetic relatedness between common and rare arthropod species to regularise the parameters of the rare species, improving their predictions(Ovaskainen et al., 2025). Bayesian JSDMs can also be combined with deep learning via intake of feature embeddings from remote sensing imagery as covariates, which in one test achieved better species predictions than either a DL-SDM or JSDM alone(Tikhonov & Tikhonov, 2025), opening a way to use other capabilities of JSDMs.

Causal attribution. – Gillespie et al.(2024) have reported that their DL-SDM, which was trained on 1-m resolution airborne imagery and 652,000 plant records, correctly predicted temporal turnover after a wildfire and spatial turnover in plant communities across ecotonal-transition and forest-succession gradients. As Gillespie et al.(2024) caution, these results are only correlational but hint that DL-SDMs could in the future attribute causes to model-predicted biodiversity changes that are triggered by signals from remote and proximal sensing. What is missing, however, is a formal *causal attribution* machinery(Schrodt et al., 2025) that can prove to some level of confidence that a given action (e.g. habitat restoration or conversion) caused a model-predicted biodiversity gain or loss. Ultimately, biodiversity markets and regulatory compliance need causal evidence, not just correlational maps.

In our view therefore, the most important capability gap in DL-SDMs is our current inability to attribute specific causes to predicted change, which reduces the information content output by DL-SDMs (Criterion 3).

A scalable but limited approach is to use supervised learning of expert-labelled samples to train a model that identifies the likely driver of an image change. For example, Sims et al.(2025) human-labelled by likely driver 7000 images in which deforestation had occurred (e.g. permanent agriculture, logging, infrastructure) and used this training dataset to estimate the spatial distribution of deforestation drivers across the globe. However, this method cannot infer what would have happened in a given location without the driver. For that, counterfactuals are required(Swinfield et al., 2024). The large spatial extents covered by DL-SDMs facilitate locating *covariate-matched* counterfactual sites(Andam et al., 2008; Ren et al., 2015), which can be compared to restoration and protection sites to infer what would have happened without the action(Swinfield et al., 2024).

The broader challenge is that to improve their predictions, DL-SDMs will often use non-causal relationships between the input features and the model outputs(Shmueli, 2010). This problem is particularly prevalent if models are presented with a large number of collinear features, i.e. features that change together. The use of non-causal relationships in the fitted models has several drawbacks for the purpose of nature conservation. (1) It is crucial for acceptance by regulators and auditors that the models 'predict the right things for the right reasons'(Coeckelbergh, 2020). (2) If the aim is to use models not only for predictions, but also for inference and ecological understanding, we again prefer that they learn the true underlying relationships(Pichler & Hartig, 2023). (3) The same is true if we want to use the models to predict the outcome of interventions. As a simple example, we want to know how biodiversity would change if we left more deadwood on the ground.

As a desirable but currently speculative possibility, Zhao & Hastie(2021) have shown that causal feature selection according to the principles of causal inference (Pearl, 2009) will constrain AI models to estimate causal effects. The advantage of this approach is that for particular causal structures (in particular, in situations dominated by confounding), a causal model can be estimated without having to define a particular intervention variable *a priori*. The key input for implementing causal feature selection for a given environment is a graph of causal connections, derived from domain experts and the scientific literature at large. Thus, instead of relying on the DL-SDM to learn *de novo* that increasing coarse deadwood in forests can boost the population sizes (and thus the occupancy probabilities) of saproxylic species like fungi and beetles, this causal relationship is already known(Doerfler et al., 2018) and can be included in the graph of causal relationships, guiding the model to predict this effect when the data warrant. Causal feature selection should be complemented with explainable AI (xAI)(Amesöder et al., 2024; Molnar, 2022; Ryo et al., 2021; Zhou & Ryo, 2025) to test after model fitting whether causal learning has been successful and to understand the behaviour of the trained models, with reference to the expectations of domain experts, other stakeholders, and the literature.

Conditional on the information being included in the causal graph, which is a major assumption, we speculate that causal feature selection might also be able to infer remote effects of an environmental change, such as the remote effects of the creation or loss of a dispersal corridor, upstream watershed protection, ecological trap habitats, and pollinator-friendly habitat patches. This capability would potentially contribute to Fletcher et al.'s(2025) proposal that the ideal metric(s) should measure how different landscape configurations and interventions contribute to the persistence of biodiversity and ecosystem services.

Extrapolation to novel conditions. – Researchers have also speculated that with enough species and observations, generative-AI versions of deep-learning SDMs could learn the 'conditional distribution of species in environmental space'(Dinnage, 2024; Hirn et al., 2022) of many species at once, i.e. Hutchinson's(1959) classical "n-dimensional hypervolume niches". This is analogous to how large language models learn the statistical distributions of words in a language, which allows LLMs to generate "new data with the statistical patterns and properties of the data upon which the models were trained"(Rafiq et al., 2025). Such generative-AI SDMs share information across species to predict the distribution of species with few data points (*few-shot* learning) and potentially with no data points but with phylogenetic or trait information to link low- or no-data species to modelled species (*zero-shot* learning)(Dinnage, 2024; Harrell et al., 2025).

Because generative-AI SDMs can generate out-of-sample communities that reflect and extend real-world patterns in the training data, researchers have speculated that such models could forecast species distributions as a function of novel changes in climate, land cover, and biotic interactions, to infer ecological connectivity in data-scarce regions, to discover, observe, and extrapolate natural experiments that mimic proposed management interventions like restoration, to assess the likelihood of success of different interventions, and thus to predict spatial changes in conservation value and to identify and interventions that contribute most to biodiversity persistence(Dinnage, 2024; Hirn et al., 2022; Rafiq et al., 2025). At the individual species level, it might be possible to forecast new migratory bird stopover sites(Pettorelli et al., 2024), algal blooms(Pettorelli et al., 2024), and new populations of rare species (and to infer the causes of rarity)(Cerrejón et al., 2021).

References

Abdelwahed, H. R., Teng, M., & Rolnick, D. (2024). *Predicting Species Occurrence Patterns from Partial Observations* (arXiv:2403.18028). arXiv. https://doi.org/10.48550/arXiv.2403.18028

Abdelwahed, H. R., Teng, M., Zbinden, R., Pollock, L., Larochelle, H., Tuia, D., & Rolnick, D. (2026). CISO: Species distribution modelling Conditioned on Incomplete Species

Observations. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.70238

Amesöder, C., Hartig, F., & Pichler, M. (2024). ‘cito’: An R package for training neural networks using ‘torch’. *Ecography*, *2024*(6), e07143. https://doi.org/10.1111/ecog.07143

Andam, K. S., Ferraro, P. J., Pfaff, A., Sanchez-Azofeifa, G. A., & Robalino, J. A. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences*, *105*(42), 16089–16094. https://doi.org/10.1073/pnas.0800437105

Baggström, A., Goodsell, R., Van Dijk, L., Iwaszkiewicz-Eggebrecht, E., Miraldo, A., Tack, A. J. M., & Andermann, T. (2025). The utility of combining deep learning with metabarcoding to model biodiversity dynamics at a national scale. *Ecological Informatics*, *90*, 103318. https://doi.org/10.1016/j.ecoinf.2025.103318

Bourel, B., Joly, A., Servajean, M., Bettinger, S., Sanabria-Fernández, J. A., & Mouillot, D. (2025). From Presence-Only to Abundance Species Distribution Models Using Transfer Learning. *Ecology Letters*, *28*(7), e70177. https://doi.org/10.1111/ele.70177

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., & Kohli, P. (2025). *AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data* (arXiv:2507.22291). arXiv. https://doi.org/10.48550/arXiv.2507.22291

Cerrejón, C., Valeria, O., Marchand, P., Caners, R. T., & Fenton, N. J. (2021). No place to hide: Rare plant detection through remote sensing. *Diversity and Distributions*, *27*(6), 948–961. https://doi.org/10.1111/ddi.13244

Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

Cole, E., Deneu, B., Lorieul, T., Servajean, M., Botella, C., Morris, D., Jojic, N., Bonnet, P., & Joly, A. (2020). *The GeoLifeCLEF 2020 Dataset* (arXiv:2004.04192). arXiv. https://doi.org/10.48550/arXiv.2004.04192

CSRD. (2022). DIRECTIVE (EU) 2022/2464 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting. *Official Journal of the European Union*. http://data.europa.eu/eli/dir/2022/2464/oj

Davis, C. L., Bai, Y., Chen, D., Robinson, O., Ruiz-Gutierrez, V., Gomes, C. P., & Fink, D. (2023). Deep learning with citizen science data enables estimation of species diversity and composition at continental extents. *Ecology*, *104*(12), e4175. https://doi.org/10.1002/ecy.4175

Deneu, B., Joly, A., Bonnet, P., Servajean, M., & Munoz, F. (2022). Very High Resolution Species Distribution Modeling Based on Remote Sensing Imagery: How to Capture Fine-Grained and Large-Scale Vegetation Ecology With Convolutional Neural Networks? *Frontiers in Plant Science*, *13*, 839279. https://doi.org/10.3389/fpls.2022.839279

Deneu, B., Servajean, M., Bonnet, P., Botella, C., Munoz, F., & Joly, A. (2021). Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLOS Computational Biology*, *17*(4), e1008856. https://doi.org/10.1371/journal.pcbi.1008856

Dinnage, R. (2024). *NicheFlow: Towards a foundation model for Species Distribution Modelling*. BioRXiv. https://doi.org/10.1101/2024.10.15.618541

Doerfler, I., Gossner, M. M., Müller, J., Seibold, S., & Weisser, W. W. (2018). Deadwood enrichment combining integrative and segregative conservation elements enhances biodiversity of multiple taxa in managed forests. *Biological Conservation*, *228*, 70–78. https://doi.org/10.1016/j.biocon.2018.10.013

Dollinger, J., Brun, P., Sainte Fare Garnot, V., & Wegner, J. D. (2024). Sat-SINR: High-Resolution Species Distribution Models Through Satellite Imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-2–2024*, 41–48. ISPRS TC II Mid-term Symposium “The Role of Photogrammetry for a Sustainable World” - 11–14 June 2024, Las Vegas, Nevada, USA. https://doi.org/10.5194/isprs-annals-X-2-2024-41-2024

Estopinan, J., Servajean, M., Bonnet, P., Munoz, F., & Joly, A. (2022). Deep Species Distribution Modeling From Sentinel-2 Image Time-Series: A Global Scale Analysis on the Orchid Family. *Frontiers in Plant Science*, *13*, 839327. https://doi.org/10.3389/fpls.2022.839327

Fletcher, R. J., Green, R. E., Bladon, E. K., Atkinson, P. W., Phalan, B. T., Williams, D., Visconti, P., & Balmford, A. (2025). Beyond Species Richness for Biological Conservation. *Conservation Letters*, *18*(4), e13124. https://doi.org/10.1111/conl.13124

Gillespie, L. E., Ruffley, M., & Exposito-Alonso, M. (2024). Deep learning models map rapid plant species changes from citizen science and remote sensing data. *Proceedings of the National Academy of Sciences*, *121*(37), e2318296121. https://doi.org/10.1073/pnas.2318296121

Giménez-Romero, À., Ferchichi, D., Moreno-Spiegelberg, P., Sintes, T., & Matías, M. A. (2024). *Mapping the distribution of seagrass meadows from space with deep convolutional neural networks* (2024.03.21.586047; p. 2024.03.21.586047). bioRxiv. https://doi.org/10.1101/2024.03.21.586047

Harrell, L., Kaeser-Chen, C., Ayan, B. K., Anderson, K., Conserva, M., Kleeman, E., Neumann, M., Overlan, M., Chapman, M., & Purves, D. (2025). *Heterogeneous graph neural networks for species distribution modeling* (arXiv:2503.11900). arXiv. https://doi.org/10.48550/arXiv.2503.11900

Hartig, F., Abrego, N., Bush, A., Chase, J. M., Guillera-Arroita, G., Leibold, M. A., Ovaskainen, O., Pellissier, L., Pichler, M., Poggiato, G., Pollock, L., Si-Moussi, S., Thuiller, W., Viana, D. S., Warton, D. I., Zurell, D., & Yu, D. W. (2024). Novel community data in ecology – properties and prospects. *Trends in Ecology & Evolution*, *39*(3), 280–293. https://doi.org/10.1016/j.tree.2023.09.017

Haucke, T., Harrell, L., Shen, Y., Klein, L., Rolnick, D., Gillespie, L., & Beery, S. (2025). *Deep Multi-modal Species Occupancy Modeling* (p. 2025.09.06.674602). bioRxiv. https://doi.org/10.1101/2025.09.06.674602

Hirn, J., García, J. E., Montesinos-Navarro, A., Sánchez-Martín, R., Sanz, V., & Verdú, M. (2022). A deep Generative Artificial Intelligence system to predict species coexistence patterns. *Methods in Ecology and Evolution*, *13*(5), 1052–1061. https://doi.org/10.1111/2041-210X.13827

Hu, Y., Si-Moussi, S., & Thuiller, W. (2025). Introduction to deep learning methods for multi-species predictions. *Methods in Ecology and Evolution*, *16*(1), 228–246. https://doi.org/10.1111/2041-210X.14466

Hutchinson, G. E. (1959). Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *The American Naturalist*, *93*(870), 145–159.

Kellenberger, B., Winner, K., & Jetz, W. (2026). The Performance and Potential of Deep Learning for Predicting Species Distributions. *Global Ecology and Biogeography*, *35*(1), e70184. https://doi.org/10.1111/geb.70184

Leblanc, C., Picek, L., Deneu, B., Bonnet, P., Servajean, M., Palard, R., & Joly, A. (2025). Mapping Biodiversity at Very-High Resolution in Europe. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2340–2349. https://doi.org/10.1109/CVPRW67362.2025.00221

Li, Y., Devenish, C., Tosa, M. I., Luo, M., Bell, D. M., Lesmeister, D. B., Greenfield, P., Pichler, M., Levi, T., & Yu, D. W. (2024). Combining environmental DNA and remote sensing for efficient, fine-scale mapping of arthropod biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1904), 20230123. https://doi.org/10.1098/rstb.2023.0123

Mackenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (Second edition). Christoph Molnar.

Ovaskainen, O., Winter, S., Tikhonov, G., Abrego, N., Anslan, S., deWaard, J. R., deWaard, S. L., Fisher, B. L., Furneaux, B., Hardwick, B., Kerdraon, D., Pentinsaari, M., Raharinjanahary, D., Rajoelison, E. T., Ratnasingham, S., Somervuo, P., Sones, J. E., Zakharov, E. V., Hebert, P. D. N., … Dunson, D. (2025). Common to rare transfer learning (CORAL) enables inference and prediction for a quarter million rare Malagasy arthropods. *Nature Methods*, *22*(10), 2074–2082. https://doi.org/10.1038/s41592-025-02823-y

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Peng, J., Li, J., Krause, J. R., Lyons, M. B., Murray, N. J., Schill, S. R., Roelfsema, C. M., & Asner, G. P. (2026). Global high-resolution mapping of seagrass to support conservation. *Nature*, 1–7. https://doi.org/10.1038/s41586-026-10704-3

Pettorelli, N., Williams, J., Schulte to Bühne, H., & Crowson, M. (2024). Deep learning and satellite remote sensing for biodiversity monitoring and conservation. *Remote Sensing in Ecology and Conservation*, *11*(2), 123–132. https://doi.org/10.1002/rse2.415

Picard, J., Nungi-Pambu Dembi, M. M., Barbier, N., Cornu, G., Couteron, P., Forni, E., Gibbon, G., Lim, F., Ploton, P., Pouteau, R., Tresson, P., van Loon, T., Viennois, G., & Réjou-Méchain, M. (2024). Combining satellite and field data reveals Congo's forest types structure, functioning and composition. *Remote Sensing in Ecology and Conservation*, *11*(2), 200–220. https://doi.org/10.1002/rse2.419

Pichler, M., & Hartig, F. (2023). *Can predictive models be used for causal inference?* (arXiv:2306.10551). arXiv. https://doi.org/10.48550/arXiv.2306.10551

Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., & Thuiller, W. (2021). On the Interpretations of Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, *36*(5), 391–401. https://doi.org/10.1016/j.tree.2021.01.002

Prudic, K. L., McFarland, K. P., Oliver, J. C., Hutchinson, R. A., Long, E. C., Kerr, J. T., & Larrivée, M. (2017). eButterfly: Leveraging Massive Online Citizen Science for Butterfly Conservation. *Insects*, *8*(2), Article 2. https://doi.org/10.3390/insects8020053

Rafiq, K., Beery, S., Palmer, M. S., Harchaoui, Z., & Abrahms, B. (2025). Generative AI as a tool to accelerate the field of ecology. *Nature Ecology & Evolution*, *9*(3), 378–385. https://doi.org/10.1038/s41559-024-02623-1

Ren, G., Young, S. S., Wang, L., Wang, W., Long, Y., Wu, R., Li, J., Zhu, J., & Yu, D. W. (2015). Effectiveness of China's National Forest Protection Program and nature reserves:

Deforestation and Protected Areas in China. *Conservation Biology*, *29*(5), 1368–1377. https://doi.org/10.1111/cobi.12561

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., & Hsiang, S. (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-24638-z

Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., & Hartig, F. (2021). Explainable artificial intelligence enhances the ecological interpretability of black‐box species distribution models. *Ecography*, *44*(2), 199–205. https://doi.org/10.1111/ecog.05360

Schrodt, F., Beck, M., Estopinan, J., Bowler, D. E., Fontaine, C., Gaüzère, P., Goury, R., Grenié, M., Martins, I. S., Morueta-Holme, N., Santini, L., Hedde, M., Martin, G., Porcher, E., Si-Moussi, S., Tzivanopoulos, M., Vernham, G., Violle, C., & Thuiller, W. (2025). Advancing causal inference in ecology: Pathways for biodiversity change detection and attribution. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.70131

Seo, E., Hutchinson, R. A., Fu, X., Li, C., Hallman, T. A., Kilbride, J., & Robinson, W. D. (2021). *StatEcoNet: Statistical Ecology Neural Networks for Species Distribution Modeling* (arXiv:2102.08534). arXiv. https://doi.org/10.48550/arXiv.2102.08534

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3). https://doi.org/10.1214/10-STS330

Sims, M. J., Stanimirova, R., Raichuk, A., Neumann, M., Richter, J., Follett, F., MacCarthy, J., Lister, K., Randle, C., Sloat, L., Esipova, E., Jupiter, J., Stanton, C., Morris, D., Melhart Slay, C., Purves, D., & Harris, N. (2025). Global drivers of forest loss at 1 km resolution. *Environmental Research Letters*, *20*(7), 074027. https://doi.org/10.1088/1748-9326/add606

Stewart, A. J., Lehmann, N., Corley, I. A., Wang, Y., Chang, Y.-C., Braham, N. A. A., Sehgal, S., Robinson, C., & Banerjee, A. (2023). SSL4EO-L: Datasets and Foundation Models for Landsat Imagery. *Advances in Neural Information Processing Systems*, *36*, 59787–59807.

Stomberg, T. T., Leonhardt, J., Weber, I., & Roscher, R. (2023). Recognizing protected and anthropogenic patterns in landscapes using interpretable machine learning and satellite imagery. *Frontiers in Artificial Intelligence*, *6*, 1278118. https://doi.org/10.3389/frai.2023.1278118

Swinfield, T., Shrikanth, S., Bull, J. W., Madhavapeddy, A., & zu Ermgassen, S. O. S. E. (2024). Nature-based credit markets at a crossroads. *Nature Sustainability*, *7*, 1217–1220. https://doi.org/10.1038/s41893-024-01403-w

Teng, M., Elmustafa, A., Akera, B., Bengio, Y., Abdelwahed, H. R., Larochelle, H., & Rolnick, D. (2023). *SatBird: Bird Species Distribution Modeling with Remote Sensing and Citizen Science Data* (arXiv:2311.00936). arXiv. https://doi.org/10.48550/arXiv.2311.00936

Tiel, N. van, Zbinden, R., Dalsasso, E., Kellenberger, B., Pellissier, L., & Tuia, D. (2025). Multi-scale and Multimodal Species Distribution Modeling. In A. Del Bue, C. Canton, J. Pont-Tuset, & T. Tommasi (Eds.), *Computer Vision – ECCV 2024 Workshops* (pp. 151–159). Springer Nature Switzerland.

Tikhonov, G., & Tikhonov, D. (2025). Synthesizing Joint and Deep Species Distribution Modeling to Enhance Spatial Prediction of Plant Communities at Continental Scale. In G. Faggioli, N. Ferro, P. Rosso, & D. Spina (Eds.), *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025)* (Vol. 4038). https://ceur-ws.org/Vol-4038/paper_261.pdf

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in ecological surveys: Estimating false-negative error rates. *Ecological Applications*, *13*(6), 1790–1801. https://doi.org/10.1890/02-5078

Zbinden, R., Tiel, N. van, Kellenberger, B., Hughes, L., & Tuia, D. (2024). On the selection and effectiveness of pseudo-absences for species distribution modeling with deep learning. *Ecological Informatics*, *81*, 102623. https://doi.org/10.1016/j.ecoinf.2024.102623

Zhao, Q., & Hastie, T. (2021). Causal Interpretations of Black-Box Models. *Journal of Business & Economic Statistics*, *39*(1), 272–281. https://doi.org/10.1080/07350015.2019.1624293

Zhou, Y., & Ryo, M. (2025). *From Images to Insights: Explainable Biodiversity Monitoring with Plain Language Habitat Explanations* (arXiv:2506.10559). arXiv. https://doi.org/10.48550/arXiv.2506.10559

Table S2. Recent DL-SDM studies. Twenty studies since 2021 that have combined deep learning with remote-sensing-derived environmental covariates (sometimes supplemented with tabular soil, climate, and land-cover covariates) to predict species distributions or summary biodiversity metrics. Studies are sorted by taxon.

Model outputs	Taxon	Citation	Environmental measures	Source and brief description of the species or class being modelled	The deep learning model approach	Model outputs and performance metrics, including comparisons to non-deep-learning models
SDMs	Plants	Deneu, B., et al. (2021). Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLOS Computational Biology*, 17(4), e1008856.	77 environmental raster variables (climatic, soil, elevation, land cover classes) for 4096 (= 64 X 64) pixels surrounding each sampling point, representing the local landscape.	GBIF dataset, filtered to 97,683 presence-only observations of 4,520 plant species across France.	A Convolutional Neural Net (CNN) was used to reduce the dimensionality of the 64 X 64 X 77 environmental tensor around each sampling point. The output vector was then converted to a probability of presence at each point for all 4,520 species.	Model performance measured using 'top-k accuracy'. The CNN model was better than RandomForest, Boosted Regression Trees, and a DNN that only used the environmental values at the sampling point (a 'punctual' model)). Ablation tests confirmed that the CNN-DNN model's ability to capture the spatial structure of the environment explained its superior performance.
SDMs	Plants	Hu, Y., et al. (2025). Introduction to deep learning methods for multi-species predictions. *Methods in Ecology and Evolution* 16(1), 228-246.	VHR (20 cm) Infra-Red (IR), Red, and Green airborne imagery used to create 100 X 100 m image patches around each vegetation plot. Source is IGN in France. Bioclimatic, pedologic, topographic, and vegetation phenology variables from public databases.	CBNA dataset, which consists of 130,582 vegetation surveys for 2522 species across the French Alps	Four models were compared: a multilayer perceptron (MLP) for tabular environmental data, a CNN and a Vision Transformer (ViT) for images, and a multimodal model combining both data types.	All models achieved True Skill Statistics (TSS) of ~70% (64.26 to 76.87%). The multimodal model performed best (>71%), and the ViT performed worst (<67%). Non-deep-learning models were not tested.
SDMs	Plants	Gillespie, L. E., et al. (2024). Deep learning models map rapid plant species changes from citizen science and remote sensing data. *Proceedings of the National Academy of Sciences*, 121(37), e2318296121.	High resolution RGB and IR airborne imagery (1 m resolution) used to create 256 X 256 m image patches centred on species observations. Source is the NAIP in the US. 1 km resolution bioclimatic variables also included.	Primarily iNaturalist dataset, filtered to 652,027 presence-only observations of 2,221 plant species across California, USA	A multimodal model called Deepbiosphere, which combines a pre-trained CNN (TResNet) for processing imagery and an MLP for processing climate data. Training used a sampling-bias-aware loss function to downweight species absence information, since many apparent absences in presence-only citizen-science datasets are false negatives.	Deepbiosphere (median AUC 0.95) outperformed the individual CNN and MLP models and the classical SDMs MaxEnt and Random Forest (median AUC 0.88). It was also able to create fine-scale range maps, ecotone edges, and detect rapid (over 2 years) anthropogenic effects on plant communities, including deforestation, habitat fragmentation, and severe wildfire.
SDMs	Plants	Leblanc, C., Picek, L., Deneu, B., Bonnet, P., Servajean, M., Palard, R., & Joly, A. (2025). Mapping Biodiversity at Very-High Resolution in Europe. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2340–2349. https://doi.org/10.1109/CVPRW67362.2025.00221	Sentinel-2 RGB and NIR imagery with 10 m resolution and Landsat time series of RGB, NIR, and SWIR1+2 imagery with 30 m resolution used to create 50 X 50 m image patches centred on species observations. A 20-year climate time series also included.	GeoPlant dataset, containing over 11,000 plant species from 5 million GBIF presence-only records in Europe and 90,000 European Vegetation Archive (EVA) presence-absence surveys.	Three CNN models (ResNet-6) were used to encode the three input modalities (Sentinel-2 images, Landsat time series, and climate time series). The outputs were concatenated and passed to a classifier that calculates species presence probabilities. The predicted occupancies per species in each of the 5.5 billion 50x50m grid cells were summarised into maps of seven one-dimensional biodiversity indicators and were also processed by Pl@ntBERT, a transformer-based large language model, to classify grid cells into EUNIS habitats.	The model with all three input modalities successfully generated maps for 5,558 species with an overall AUC score of 0.931 but an F-score of only 0.338, where F-scores are more appropriate for imbalanced datasets because they count the avoidance of both false-positive and false-negative error. These scores were better than those from the models using only one input modality. Biodiversity indicator maps include species richness and probability that at least 1 threatened species is present. EUNIS habitat prediction accuracies ranged from 42.8% to 76.3%, depending on habitat level and number of species used.

SDMs	Plants	Tiel, N. van, Zbinden, R., Dalsasso, E., Kellenberger, B., Pellissier, L., & Tuia, D. (2025). Multi-scale and Multimodal Species Distribution Modeling. In A. Del Bue, C. Canton, J. Pont-Tuset, & T. Tommasi (Eds.), Computer Vision – ECCV 2024 Workshops (pp. 151–159). Springer Nature Switzerland.	Sentinel-2 RGB and NIR imagery at 10 m resolution and 19 bioclimatic rasters at ~600 m resolution. The study tested the effect of using different patch sizes. Individual satellite pixels were merged into image patches centred over species observations at 0.06, 0.35, and 1.33 sq. km sizes. Bioclimatic pixels were merged into patches from 0.36 to 225 sq. km.	GeoLifeCLEF 2023 dataset for 2,173 plant species, containing 5 million presence-only observations across Europe and 7,438 presence-absence survey sites in France and UK.	Multimodal, multiscale CNN architecture to incorporate both bioclimatic and satellite imagery, with different patch sizes, after which the outputs were concatenated and passed to a classifier that calculates species presence probabilities. Uniscale/unimodal model variants were tested to quantify the effect of using multiple scales and modalities.	Predictive AUC scores were highest for the multimodal models, followed closely by the bioclimatic-only model. Predictive F-scores were highest for the multimodal, multiscale model, potentially dependent on species and numbers of observations.
SDMs	Plants	Dollinger, J., Brun, P., Sainte Fare Garnot, V., & Wegner, J. D. (2024). Sat-SINR: High-Resolution Species Distribution Models Through Satellite Imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-2–2024, 41–48.	Sentinel-2 RGB and NIR imagery at 10 m resolution, median value across a whole year, used to create 1.28 X 1.28 km image patches centred on species observations. Bioclimatic and location data at a resolution of 1 km also included.	GeoLifeCLEF 2023 dataset. For training, the dataset is filtered to contain 2 million observations of about 5,600 plant species, primarily across western Europe. For testing, the dataset is filtered to 6,000 presence-absence surveys in France and UK, containing 2000 species.	Sat-SINR model. A multimodal CNN was used to encode the satellite imagery, and the 1-dimensional outputs were concatenated with the climate data and passed to a classifier that calculates species presence probabilities. The study tested early, mid, and late fusion strategies for combining the data modalities.	The multimodal models performed better in predictive AUC and F-scores than the unimodal model using only bioclimatic and location data. Altering fusion method has small effects, with late fusion achieving the best F-score.
SDMs	Plants	Tikhonov, G., & Tikhonov, D. (2025). Synthesizing Joint and Deep Species Distribution Modeling to Enhance Spatial Prediction of Plant Communities at Continental Scale. In G. Faggioli, N. Ferro, P. Rosso, & D. Spina (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025) (Vol. 4038). https://ceur-ws.org/Vol-4038/paper_261.pdf	Sentinel-2 RGB+NIR imagery and WorldCover classification (640×640 m tiles at 10 m resolution); Landsat time series (2000–2017, quarterly, 30 m); CHELSA monthly temperature and precipitation, snow cover duration (2000-2017, quarterly); soil variables; elevation; 7 land cover classes	GeoLifeCLEF 2025 dataset. Plant communities across Europe; 11,255 anonymised species (5,016 from presence–absence surveys, 9,709 from presence-only occurrences); 89K PA survey records 2017–2021, 43 countries across 9 biogeographic regions; 5m PO records from GBIF. Highly heterogeneous survey effort and design by country.	Stage 1: multimodal DNN (Prithvi-EO-2.0-300M & ResNet18 for imagery + MLP for tabular data) trained as a standard deep SDM. Stage 2: PCA-reduced DNN feature embeddings passed as covariates into Hmsc-HPC, a JSDM, with spatial random effects.	Species occurrence probabilities for 11,255 taxa; evaluated by F1 score. Best model (DNN -> 228 PCs -> HMSC + spatial random effects): F1_test = 0.217. Ensemble of 15 variant models: F1_test = 0.222. Ranked 2nd overall in competition. DNN + HMSC performed better than either alone.
SDMs	Orchids	Estopinan, J., et al. (2022). Deep Species Distribution Modeling From Sentinel-2 Image Time-Series: A Global Scale Analysis on the Orchid Family. *Frontiers in Plant Science*, 13, 839327.	Sentinel-2 RGB and IR imagery used to create a 12-month time series of 10-m resolution, 640 x 640 m image patches over each observation location. The time series is used to capture seasonal changes in vegetation phenology, weather, and human activity.	GBIF dataset, filtered to 999,407 presence-only observations of 14,148 orchid species (Orchidaceae) across the globe	An extended Inception v3 CNN architecture modified to accept the RGB+IR monthly time series as input. Training used a class-imblance-aware loss function optimised to increase performance on rare species. To test the importance of the time series, model variants were also run on shuffled, averaged, and single-month datasets.	The model using the full, ordered time series outperformed the models trained with the shuffled, averaged, and single-month datasets (top-30 accuracy averaged across all species of 0.286 vs. 0.149-0.216). The performance gain from use of the time series information was greatest for rare species and in regions with high species diversity. Non-deep-learning models not tested.
SDMs	Birds	Davis, C. L., et al. (2023). Deep learning with citizen science data enables estimation of species diversity and composition at continental extents. *Ecology*, 104(12), e4175.	64 environmental covariates from remote sensing data, including land-use/land-cover (LULC), topography, and hydrology layers derived from the NASA MODIS satellite and nighttime reflectance measures from the spaceborne NOAA VIIRS sensors, plus 8 checklist-specific covariates to account for variation in detection rates.	eBird dataset, over 9 million presence-absence checklists of 500 North American bird species	The study applied the DMVP-DRNets model (Deep Reasoning Network implementation of the Deep Multivariate Probit Model), which is a deep-learning version of a joint species distribution model (JSDM). An encoder-decoder framework with three multilayer perceptrons (MLPs) was used to fit the model.	The primary model output was monthly individual species distributions across North America at a spatial resolution of 2.9 km, which were used to dynamically track areas of high species richness and their overlap with existing Bird Conservation Regions. DMVP-DRNets outperformed a Bayesian JSDM over eight datasets (birds, butterflies, plants, trees, vegetation).

SDMs	Birds	Teng, M., et al. (2023). SatBird: Bird Species Distribution Modeling with Remote Sensing and Citizen Science Data. *arXiv:2311.00936*.	Sentinel-2 satellite RGB and Near-Infrared (NIR) imagery at 10 m resolution, 19 bioclimatic rasters from WorldClim, and 8 pedologic rasters from SoilGrids.	SatBird dataset, filtered from the eBird dataset to 122,593 summer and 53,361 winter presence-absence checklists of 670 bird species in the USA	The study tested four different pre-trained CNN foundation models (ResNet-18, MOSAIKS, SATLAS, and SatMAE) as encoders to extract features from the satellite imagery to use as predictors. The study then applied a single-layer MLP to predict species distributions.	ResNet-18 models that used satellite imagery plus the bioclimatic and soil rasters performed better than the other encoders and better than the baseline Boosted Regression Trees using the bioclimatic and pedological variables only, especially using 'top-K' accuracy, which is the proportion of observed species found in the model's top 'k' predictions ranked by probability.
SDMs	Birds	Seo, E., Hutchinson, R. A., Fu, X., Li, C., Hallman, T. A., Kilbride, J., & Robinson, W. D. (2021). StatEcoNet: Statistical Ecology Neural Networks for Species Distribution Modeling *arXiv:2102.08534*.	Thirty-four spectral indices engineered from Landsat satellite image composites, including Tasseled Cap indices (brightness, greenness, wetness, angle) and topographic features (aspect, elevation, slope) at different spatial scales.	Five bird species selected from an expert-conducted point count survey in Oregon, USA. At each sampling point, 3-8 replicate surveys were carried out, which allows occupancy correction to be applied.	StatEcoNet, a framework that integrates two separate neural networks into a classic statistical occupancy model. One network models the occupancy probability from site features (ecological model), and the other models the detection probability from survey features (observation model).	On both simulated and real bird data, StatEcoNet outperformed linear, machine learning, and other neural network models. For simulated nonlinear data, StatEcoNet produced more accurate probability estimates, higher AUC values, and correctly identified the most important covariates. For real bird data, it achieved a higher Area Under the Precision-Recall Curve than other occupancy models.
SDMs	Mammals	Haucke, T., et al. (2025). Deep Multi-modal Species Occupancy Modeling. *bioRxiv*:2025.09.06.674602.	~100 environmental covariates, including bioclimatic and soil vectors, satellite imagery embeddings extracted by Alpha Earth Foundations, and camera trap imagery embeddings extracted by DINOv2 FiT-B/14.	16 North American mammal species with high prevalence in camera trap data from Wildlife Insights platform.	Deep learning was only used to extract embeddings from satellite and camera trap imagery, which were used as environmental covariates in the Bayesian occupancy model Biolith. xAI was used to identify camera trap images that have strong predictive power for presence and absence of individual species. These images were presented to the investigator for qualitative interpretation.	Since all non-detections are suspect in an occupancy modelling framework, predictive performance was assessed by summing the log-averaged likelihoods of held-out observations under the posterior predictive distribution. Satellite embeddings contributed the most to predictive performance, followed closely by camera trap embeddings. Standard bioclimatic and soil data contributed little to predictive performance.
SDMs	Fish	Bourel, B., et al. (2025). From Presence-Only to Abundance Species Distribution Models Using Transfer Learning. *Ecology Letters*, 28(7), e70177.	Sentinel-2 RGB and NIR imagery at 10 m resolution, plus geographic coordinates, bathymetry, chlorophyll concentration, sea surface temperature, sea water velocity and salinity, and sea floor temperature and substrate. Rasters windows constructed at 3 X 3, 30 X 30, and 63 X 63 km.	GBIF dataset, filtered to 62,240 presence-only occurrences of 181 fish species up to 20 km from the coast of the Mediterranean Sea, and the Reef life Survey count dataset in GBIF of 406 transect surveys for 47 species in 217 sites.	A CNN-SDM with a ResNet-50 architecture modified to take all 37 environmental dimensions. The key innovation is transfer learning from a model trained on the large presence-only dataset, after which the final layer is de novo trained on the smaller abundance dataset. This model was compared with RandomForest.	Using the D-squared score function (D2log) and the Spearman Rank correlation as performance metrics, the transfer learning CNN model showed better abundance prediction over both the RandomForest and CNN model without transfer learning, especially for the rare species.
SDMs	Arthropods	Li, Y., et al. (2024). Combining environmental DNA and remote sensing for efficient, fine-scale mapping of arthropod biodiversity. *Philosophical Transactions of the Royal Society B*, 379(1904), 20230123.	29 environmental covariates, including raw Landsat bands and indices calculated from Landsat data, canopy structural indices from airborne lidar data, and anthropogenic and topographic layers calculated from publicly available GIS layers, all mapped at 30 m resolution across a 225 sq. km working forest landscape in Oregon, USA.	An arthropod community dataset of 190 arthropod species was generated from environmental DNA (eDNA): 121 shotgun-sequenced Malaise-trap samples, with species detected via mapping of raw reads to a reference set of DNA barcodes.	The study applied the sjSDM model (sparse jSDM). Two submodels were fitted using deep neural nets. One fitted the environmental covariates to model environmental filtering, and the other fitted a trend surface to model local dispersal.	The fitted model achieved median and mean predictive AUC values of 0.86. Explanatory AI analysis showed that 18 of 29 covariates were the most important for at least one species, showing that the model made good use of the predictors. 76 species had individual predictive AUCs ≥0.7 (mean 0.83), and these species distributions were reduced to single-dimensional maps showing species richness, site irreplaceability, and species compositional differences (via t-sne values).

SDMs	Plants and animals	Deneu, B., et al. (2022). Very High Resolution Species Distribution Modeling Based on Remote Sensing Imagery: How to Capture Fine-Grained and Large-Scale Vegetation Ecology With Convolutional Neural Networks? *Frontiers in Plant Science*, 13, 839279.	High and Very High Resolution (VHR) RGB and NIR airborne imagery used to create 256 X 256 m image patches centred on species observations. Sources are the NAIP (1 m resolution) in the US and IGN (0.2, 0.5 m) in France. Lower-resolution land cover classes and elevation layers also included.	GeoLifeCLEF 2020 dataset, which consists of 1,921,123 presence-only observations of 31,435 mostly plant and animal species across France and the US	A CNN was used to reduce the dimensionality of each image patch to a 2048-element output vector, which was then converted to a probability of presence at each point for each species.	Model performance measured using 'top-30 accuracy' was better than a RandomForest model that used bioclimatic and soil rasters at 1 km resolution. The CNN model trained only on the imagery consistently performed better than the RF model, especially for rarer species.
SDMs	Birds and Butterflies	Abdelwahed, H. R., et al. (2024). Predicting Species Occurrence Patterns from Partial Observations. *arXiv:2403.18028*.	Sentinel-2 RGB and NIR imagery with 10 m resolution used to create 5 sq. km image patches centred on species observations. 19 bioclimatic rasters at 1 km resolution and 8 pedologic rasters also included.	The SatButterfly v2 dataset, which is derived from the eButterfly dataset and comprises a filtered set of 6,684 presence-absence checklists of 172 species in the US, in locations that overlap with SatBird, with 670 bird species.	The R-Tran model, which uses ResNet-18 CNNs to extract features from the satellite, climate, and soil rasters. The outputs are then concatenated with a variable number of known and co-located bird observations. From this, a transformer encoder model is used to model interactions between the birds and environmental features with butterfly species, to predict unknown butterfly or bird species conditional on known observations of (other) bird species.	R-Tran was tested on its ability to predict non-songbirds given songbird and environmental observations and to predict butterflies given all bird and environmental observations. Both tests showed small improvements over non-conditional-prediction models.
Summary metrics	Six forest types	Picard, J., et al. (2024). Combining satellite and field data reveals Congo's forest types structure, functioning and composition. Remote Sensing in Ecology and Conservation, 11, 200–220.	Sentinel-2 imagery data at 10 m resolution covering the 18 500 sq. km study area in the north of the Republic of Congo. The 13 spectral bands were reduced to three principal components and converted to false RGB images.	535 1-ha polygons from VHR Pleiades satellite stereo imagery at 0.5 m resolution, supplemented with 145 1-ha polygons from Planet imagery at 5 m resolution of forest. The polygons were labelled by a single expert to one of 7 classes: monodominant Gilbertiodendron forests, dense forests, Marantaceae forests, open Marantaceae forests, secondary forests, swamp forests, and roads.	A CNN (EfficientNet-B3 with weights from Noisy-Student training) and a Vision transformer pre-trained on ImageNet were each used to train models to classify each 10 X 10 m pixel to each of the 7 classes. Post-training probabilities were averaged across all models, and each pixel was assigned to the highest probability class.	Overall classification accuracy was 83%, with most classification errors occurring between similar forest types (e.g. dense versus Marantaceae, and Marantaceae versus open Marantaceae). The six forest types showed significant differences in independent tree-plot measures of forest structure, water table depth, soil types, and tree species composition.
Summary metrics	Protected versus non-protected areas	Stomberg, T. T., et al. (2023). Recognizing protected and anthropogenic patterns in landscapes using interpretable machine learning and satellite imagery. *Frontiers in Artificial Intelligence*, 6, 1278118.	Sentinel-2 satellite imagery data at 10 m resolution covering Fennoscandia, using Channels B2, B3, B4, B5, B6, B7, B8, B8A, B12, and B12. The pixels were tiled into images of size 256 X 256 pixels.	The AnthroProtect training dataset comprised images classified as belonging to category Ia, Ib, or II protected areas (using the World Database on Protected Areas) or to anthropogenic landcover classes 1 (urban fabric) or 2 (arable) (using the Copernicus CORINE LC dataset).	Protected and Anthropogenic images were "CutMixed" by pasting a stripe from a random sample to a random edge, adding 0-50% to the original image size, and adjusting each image's label to reflect the proportions of the Protected and Anthropogenic images. A CNN model was trained to predict a continuous value between 0 (anthropogenic) to 1 (protected). Prediction scores were harmonised during training to remove the effect of contrast (or lack thereof) with surrounding landscape.	The model successfully learned to distinguish between protected and non-protected areas, achieving a low root mean squared error when classifying CutMixed images with a continuous score.

Summary metrics	Seagrass meadows	Giménez-Romero, À., et al. (2024). Mapping the distribution of seagrass meadows from space with deep convolutional neural networks. *bioRxiv:2024.03.21.586047*.	60 High Resolution PlanetScope-Super Dove and RapidEye imagery at ~2.9 m resolution, along the coasts of the Balearic Islands, covering 1200 sq. km, with Coastal Blue, Blue, Green I, Green, Yellow, Red, Red Edge, and NIR channels. Bathymetry data also used.	2500 sq. km habitat dataset of the coastal habitats of the Balearic Islands in the Mediterranean Sea, including Posidonia oceanica seagrass meadows as one of 28 classes, based on an extensive ground-truth dataset from side-scan sonar and photo-interpretation.	CAMELE model for semantic image segmentation, comprising 10 different CNN architectures across 40 models. Each model classified each pixel in a satellite image as seagrass or other benthic habitat. The class label with the highest frequency across all models was assigned to the pixel.	The CAMELE model correctly predicted 77.3% of the seagrass pixels in the out-of-sample dataset (Intersection over Union score), and an F-score of 72.8% across all habitat classes.
Summary metrics	Arthropod species richness	Baggström, A., et al. (2025). The utility of combining deep learning with metabarcoding to model biodiversity dynamics at a national scale. *Ecological Informatics*, 90, 103318.	25 environmental covariates, including climatic data (BioClim, ERA5), vegetation and biomass data from airborne laser scanning, Sentinel-2 derived NDVI, human impact index, edaphic and topographic layers. The ML models used only the values at the 10 m central pixel centred on the sampling location. The CNN model used a stacked raster of 128 X 128 pixels centred over the sampling location.	Arthropod metabarcoding dataset of over 33,394 Operational Taxonomic Units total from 3728 bulk samples collected weekly in spring/summer 2019 from 198 Malaise traps across Sweden, in six different habitats: forest, croplands, wetlands, grasslands, alpine, and urban areas. The target metric is species richness.	A CNN with a ResNet-18 architecture, modified for regression to predict species richness, was used to predict OTU richness from environmental image patches.	Model performance was judged using 7-fold cross validation. The CNN did not outperform Random Forest or XGBoost on quantitative metrics (RMSE, MAE, R-squared). However, the fine-scale spatial patterns predicted by the CNN matched an independent dataset of land-use classes, whereas the ML models did not.