PUBLICATION INFORMATION

This is an accepted manuscript version of an article by Alex Bush, et.al. 2017. **Connecting Earth observation to high-throughput biodiversity data**. *Nature Ecology & Evolution, 1: 0176*.
DOI: https://doi.org/10.1038/s41559-017-0176

Connecting Earth Observation to High-Throughput Biodiversity Data

Bush, Alex[1,2,3]; Sollmann, Rahel[4]; Wilting, Andreas[5]; Bohmann, Kristine[6,7]; Cole, Beth[8]; Balzter, Heiko[8,9]; Martius, Christopher[10]; Zlinszky, András[11]; Calvignac-Spencer, Sébastien[12]; Cobbold, Christina A.[13]; Dawson, Terence P.[14]; Emerson, Brent C.[15,7]; Ferrier, Simon[3]; Gilbert, M. Thomas P.[6,16]; Herold, Martin[17]; Jones, Laurence[18]; Leendertz, Fabian H.[12]; Matthews, Louise[13]; Millington, James D.A.[14]; Olson, John R.[19]; Ovaskainen, Otso[20,21]; Raffaelli, Dave[22]; Reeve, Richard[13]; Rödel, Mark-Oliver[23]; Rodgers, Torrey W.[24]; Snape, Stewart[25]; Visseren-Hamakers, Ingrid[26]; Vogler, Alfried P.[27,28]; White, Piran C.L.[22]; Wooster, Martin J.[14]; Yu, Douglas W.[1,7,‡]

1. State Key Laboratory of Genetic Resources and Evolution, **Kunming Institute of Zoology**, Chinese Academy of Sciences, Kunming, Yunnan 650203 China
2. Environment Canada @ Canadian Rivers Institute, Department of Biology, **University of New Brunswick**, PO Box 4400, Fredericton, NB, E3B 5A3, Canada
3. **CSIRO Land and Water**, Canberra, ACT 2601, Australia
4. Department of Wildlife, Fish, & Conservation Biology, 1088 Academic Surge, One Shields Ave, **University of California Davis**, Davis, CA 95616 USA
5. **Leibniz Institute for Zoo and Wildlife Research**, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

6. EvoGenomics, Natural History Museum of Denmark, **University of Copenhagen**, 1350 Copenhagen K Denmark
7. School of Biological Sciences, **University of East Anglia**, Norwich Research Park, Norwich, Norfolk NR47TJ UK
8. Centre for Landscape and Climate Research and Leicester Institute for Space and Earth Observation (LISEO), **University of Leicester**, University Road, Leicester LE17RH UK
9. NERC National Centre for Earth Observation (NCEO) at **University of Leicester**, University Road, Leicester LE17RH, UK
10. **Center for International Forestry Research (CIFOR)**, P.O. Box 0113 BOCBD, Bogor 16000, Indonesia
11. Balaton Limnological Institute, **Centre for Ecological Research**, Hungarian Academy of Sciences, Tihany 8237 Hungary
12. **Robert Koch Institut**, Berlin 13353 Germany
13. Boyd Orr Centre for Population and Ecosystem Health, **University of Glasgow**, Glasgow G128QQ UK
14. Department of Geography, **King's College London**, Strand Campus, London WC2R2LS UK
15. **IPNA-CSIC**, La Laguna, Tenerife, Canary Islands 38206 Spain
16. NTNU University Museum, **Norwegian University of Science and Technology**, Trondheim 7491 Norway
17. Laboratory of Geo-Information Science and Remote Sensing, **Wageningen University & Research**, Wageningen, the Netherlands
18. Centre for Ecology and Hydrology, **Environment Centre Wales**, Deiniol Road, Bangor LL572UW UK
19. Division of Hydrologic Sciences, **Desert Research Institute**, Las Vegas, NV, USA
20. Department of Biosciences, **University of Helsinki**, Helsinki FI-00014 Finland

21. Centre for Biodiversity Dynamics, Department of Biology, **Norwegian University of Science and Technology**, Trondheim 7491 Norway
22. Environment Department, **University of York**, York YO105NG UK
23. Museum für Naturkunde - **Leibniz Institute for Evolution and Biodiversity Science**, Berlin 10115 Germany
24. Department of Wildland Resources, **Utah State University**, Logan, UT 84322 USA
25. **Forestry Commission**, Edinburgh, UK
26. Department of Environmental Science and Policy, **George Mason University**, Fairfax, VA 22030 USA
27. Department of Life Sciences, **Natural History Museum**, London SW75BD UK
28. Department of Life Sciences, Silwood Park Campus, **Imperial College London**, Ascot SL57PY UK

ǂ Corresponding author: dougwyu@gmail.com

Preface

There is much interest in using Earth Observation (EO) technology to track biodiversity, ecosystem functions, and ecosystem services, understandable given the fast pace of biodiversity loss. However, because most biodiversity is invisible to EO, EO-based indicators could be misleading, which can reduce the effectiveness of nature conservation and even unintentionally decrease conservation effort. We describe an approach that combines automated recording devices, high-throughput DNA sequencing, and modern ecological modelling to extract much more of the information

available in EO data. This approach is achievable now, offering efficient and near-real-time monitoring of management impacts on biodiversity and its functions and services.

Meeting the Aichi Biodiversity Targets

From Google Earth to airborne sensors, the Copernicus Sentinels, and cube satellites, Earth Observation is undergoing a rapid expansion in capacity, accessibility, resolution, and signal-to-noise ratio, resulting in a recognised shift in our capability for using remote-sensing technologies to monitor biophysical processes on land and water[1-3]. These advances are motivating calls to use Earth Observation products to manage our natural environment and to track progress toward global and national policy targets on biodiversity and ecosystem services[4-6]. Foremost among these policies are the Strategic Plan for Biodiversity and the Aichi Biodiversity Targets, which were adopted in 2010 by the Parties to the Convention on Biological Diversity (CBD) to "take effective and urgent action to halt the loss of biodiversity in order to ensure that by 2020 ecosystems are resilient and continue to provide essential services…"[7]. The United Nations Sustainable Development Goals[8] now include some of the Aichi Targets, and the 2015 Paris Agreement has reiterated the commitments of the UN Framework Convention on Climate Change to reducing emissions from deforestation and forest degradation

(REDD+) and to securing non-carbon benefits, which include biodiversity and ecosystem services[9].

However, we have struggled to track and report progress toward the Aichi Targets in a standardised and comprehensive way[10]. Although almost two-thirds of the CBD Parties have updated their National Biodiversity Strategies and Action Plans to reflect the 2010 revisions, many still do not contain measurable indicators on the state of biodiversity, let alone ecosystem services. This lack of quantification conceals the impacts of policy and management interventions on biodiversity and ecosystem functions and services[11]. The difficulty of designing indicators[12-14] has prompted an international consortium of biodiversity scientists called GEO BON (Group on Earth Observations' Biodiversity Observation Network) to propose a framework of Essential Biodiversity Variables[15], with the aim of setting minimum standards of coverage to ensure informativeness and to harmonise disparate local measures so that biodiversity and ecosystem data can be compared over space and time. The Essential Biodiversity Variables thus measure the 'state of biodiversity' at multiple levels: genetic composition, species populations, species traits, community composition, ecosystem structure, and ecosystem function[15].

Although it was originally envisioned that most of the variables (genetic to community composition) would be scaled up from "intensive *in-situ* measurements"[15] taken on the ground, such measurements are costly and difficult because they are traditionally gathered by visual and aural detection of plants and animals in the wild (preceded by months or years of observer practice) and by mass collection of organisms (followed by months of identification from morphology), so that data collection is slowed by human-caused bottlenecks in sampling and taxonomy[16].

As a result, attention is now being focused on designing 'Satellite Remote Sensing-Essential Biodiversity Variables' (SRS-EBVs) to enable cost-effective and global-scale monitoring[5,6,12]. The problem here is that only a few Earth Observation products can be mapped directly to Essential Biodiversity Variables and then to Aichi Targets, because these products primarily measure gross vegetation and landscape metrics, such as land cover and phenology[4]. For example, Pettorelli et al.[12] found only two Earth Observation products (net primary productivity and fire incidence) that could serve as Essential Biodiversity Variables for the Sahara, despite this biome's suitability for remote sensing due to its visible biodiversity hotspots, remoteness, and availability of long time series. Many of the Aichi Targets require data with species-level resolution, either because some species are direct policy targets (e.g. Target 9: "invasive species controlled or eradicated") or because species compositional data define the metric (e.g. Target 11: "protected areas are ecologically representative and conserved effectively").

Clearly, a radically new approach is required if progress towards the Aichi Targets is to be accelerated, one that is robust, widely affordable, and can record stocks and changes in biodiversity and ecosystem services consistently, continuously, and at high resolution over large geographic scales. Here, we present such an approach in a framework that exploits recent efficiency gains and analytical breakthroughs in sensors, computation, ecology, taxonomy, and genomics (**Figure 1, Box 1**).

Box 1. Inferring a Hidden Ecosystem Function from Space

Large-bodied Amazonian monkeys are responsible for a key ecosystem function: they are the primary dispersers of large seeds, which are associated with more carbon-dense tree species. Peres et al.[17] have proposed that this function boosts forest carbon storage. The idea can be tested by using Earth Observation data and public records to map human settlements and transport corridors and predict where monkey populations have declined through hunting[17,18]. We can then use on-the-ground sampling and airborne sensors to test whether forests that have had longer exposure to hunting lack monkey populations and have more low-carbon-density tree species dispersed by wind and birds. In short, by combining Earth-Observation-derived maps of human activity with empirical observations of the response of primate populations to that activity, it should be possible to map and track an ecosystem function (large-seed dispersal) that is invisible to satellites but contributes to an important ecosystem service (climate regulation).

From Point Samples to Continuous Maps

Instead of trying to map Earth Observation (EO) products directly to biodiversity, as encapsulated by SRS-EBVs[4-6,12], we propose to extract more information from EO data by interpolating biodiversity point samples to build continuous landscape maps of species distributions (**Figure 1**)[19]. Because it is species that are mapped, it then becomes possible to layer on the vast biological knowledge that we have collectively built up over decades of research, including historical distributions, phylogenetic relationships, and knowledge of species traits and interactions to infer, map, and track the distributions of ecosystem functions and services (**Box 1**). This approach, which we call here **CEOBE** (Connecting Earth Observation to Biodiversity and Ecosystems), is possible because of (1) major advances in EO sensitivity and capacity, (2) more efficient techniques to collect biodiversity data on the ground, and (3) modern community-analysis models from statistical ecology. We now review each of these advances, with additional detail in Supplementary Information.

The New Era of Earth Observation

There are ten times as many satellites in operation now as there were in the 1970s, a result of increasing sensor longevity and a six-fold increase in launches[20]. Spatial resolution has improved to less than 1 m in both optical and radar sensors. Data continuity is also being maintained, most directly by the launch of NASA's Landsat 8 in 2013, which extends and technically enhances the 40-year Landsat record of medium-resolution, multispectral surface observations[21]. Data continuity is a key factor in understanding changes in biodiversity, as threats to biodiversity impact at a range of scales and often across lengthy timespans[22].

The long-term Landsat record is being enhanced by new satellite systems and multiple sensors in a global network, a 'virtual constellation' that may help overcome problems in terrestrial monitoring from single sensors[2]. As part of the Copernicus program, the ESA Sentinel satellites are the latest addition to the global network. With six missions planned and the first three launched, the Sentinels have radar, optical sensors, radiometers, and spectrometers with different goals[23]. Sentinel-1, the radar satellite, and Sentinel-2, the superspectral high-resolution mission, are of particular interest to biodiversity monitoring, with long-term continuity of measurements, global coverage, and quick revisit times [24,25].

There have also been developments in hyperspectral sensors with EnMAP, HyspIRI, PRISMA, and FLEX imaging spectrometer missions planned[1]. In addition, airborne data collection using high-resolution 3D airborne laser scanning is complementing spectral information with structure[26]. Swarms of commercial cube satellites and the use of drones to carry sensors are additional significant steps that complement these large-scale programs (**Supplementary Note 1** "Earth Observation technology").

The increase in spatial resolution in the new sensors implies greater precision because reference measurements taken within meter-scale plots on the ground can be matched directly to meter-scale pixels[27]. This in turn improves the ability of EO to recognise spatial gradients and boundaries.

Two additional factors affect the utility of remote sensing data for understanding biodiversity change (**Supplementary Note 2** "Biodiversity and ecosystem information in EO data"): affordability and access[22]. There has been a cultural shift, with free open access on the rise. The opening of the Landsat archive in 2008 was a monumental development[28], with ESA's Copernicus program following suit. Data access also refers to the ability of users to retrieve, manipulate, and extract value from EO data. Cloud computing and toolboxes are making these processes manageable, even with large data archives.

The availability of copious EO data that have been shown in multiple studies to correlate closely with on-the-ground measures of ecosystem structure, habitat condition, and even animal communities (**Supplementary Note 2**) might suggest that remote sensors can be used directly to define environmental indicators, but we must acknowledge that we are still in the early stages of understanding how biodiversity delivers ecosystem functions and services, and how they all respond to exogenous change. Directly observing functional diversity is a partial solution but only with visible biodiversity such as vegetation[26]. Thus, the challenge is to find ways to exploit the high efficiency and information content of EO data while not falling prey to *reification fallacy* (**Box 2**), which can arise when convenient but incomplete indicators are made available[29,30]. Our institutions and reporting systems then retain the option to add and respond to new knowledge.

Box 2. The Perils of Convenient Indicators

If we rely too directly on EO data, we run the risk of *reification fallacy*, in which a mere indicator of a policy target itself ends up the target. Reification fallacy can reduce or narrow conservation effort[31] and can crowd out future discoveries[32]. For example, while remote sensing is an efficient and direct way to measure forest *cover* (Aichi Target 5:

reducing the loss rate of natural habitats), using forest cover and phenology to measure the contribution of biodiversity to carbon stocks (Target 15)[4] would ignore taxa invisible to satellites and could thus result in policymakers failing to exert the additional effort that is required to conserve saprotrophic fungal diversity, seed-dispersing mammals, and the seemingly inconsequential isopod, all of which have been implicated in boosting carbon storage[17,33,34]. More generally, land-cover class, which is a common EO-indicator, is a highly error-prone way to map and assess the complex processes supporting ecosystem services[35]. In short, convenient EO products could lead policymakers to focus only on that portion of biodiversity and ecosystem services that is directly observed by remote sensing, ignoring the rest.

High-Throughput Biodiversity Measurement

Most biodiversity, whether animal, fungal, plant, or microbial, and its many functions and services, is invisible to EO and will remain so for some time. But a growing number of efficient technologies are available for detecting and identifying biodiversity on the ground[36,37] (**Supplementary Note 3** “Biodiversity technology”). Automated bioacoustic and camera-trap recording devices (ARDs) can run continuously for weeks and accumulate thousands of records of invertebrates, birds, fish, reptiles, amphibians, and mammals, and thus allow extended sampling of large areas at low workloads[38-42]. Alternatively, high-throughput DNA sequencers can be used in metabarcoding or metagenomic pipelines to detect and identify anywhere from one to thousands of species at a time from mass-collected, bulk samples of organisms (e.g. 'biodiversity soups'[43]), or from 'environmental DNA,' which is DNA liberated into the environment in the skin, hair, mucous, saliva, sperm, eggs, exudates, faeces, urine, blood, spores, root fragments, leaves, fruit, pollen, or rotting body parts of their original owners[44,45] (**Figure 2, Supplementary Note 3**). Multiple studies have now shown that metabarcode datasets reflect high-quality, morphologically identified biodiversity datasets sufficiently closely to allow correct management decisions, given best-practice protocols and controls[46-51].

The taxonomic identities, phylogenetic affinities, functional genes[52], spectral properties (of visible vegetation[26,53,54]), and/or co-occurrence patterns[55] of the detected species can be used to parameterise process-based production functions for ecosystem services[56-58] (**Figure 1**). For instance, the species identities and biomasses of wild bees identified metagenomically from bulk samples[59] could be combined with flower-use observation data[60] and detailed vegetation classification from EO to infer the availability and nature of local pollination services. Metagenomic data matched to identified species can be particularly powerful when the impacts of species loss on ecosystem function are not random, evidence that has previously relied on intensive field sampling, e.g. in tropical freshwater[61] and marine benthic communities[62].

Statistical Modelling as the Bridge

Earth Observation technology can produce large-scale, fine-resolution maps and dense time series of a wide range of biophysical variables (**Supplementary Note 1 and 2**), but it is difficult to translate the biophysical variables into biodiversity information. In contrast, ARDs and DNA sequencing are capable of generating large amounts of biodiversity information at species- or even individual-level resolution[63,64], but only from point samples (**Supplementary Note 3**). Modern methods of statistical modelling allow us to interpolate these point samples to build continuous species maps and to estimate emergent metrics such as richness and dissimilarity[65-68], potentially also including estimates of species abundance or biomass, depending on the sampling and analytical methods used (**Supplementary Note 4** "Statistical modelling").

The three approaches with immediate potential are *Joint Species Distribution Models*[69-72] (including *Latent Variable Models*), *Community Occupancy-Detection Models*[73], and *Generalised Dissimilarity Models*[65,74] (**Figure 3**, **Supplementary Note 4**). Each approach starts with a site-by-species matrix, from data that have been collected by ARDs or been generated via metabarcoding or metagenomics (**Figure 2**, **Supplementary Note 3**), plus any existing species distribution data. If some species are not detected, repeat sampling can be used to infer missing occurrences[73]. The site-by-species matrix is then paired with a corresponding site-by-environmental-covariate matrix, generated from continuous EO data plus any relevant geographical layers, and the two datasets are combined statistically to infer the joint distributions of multiple species across entire regions (**Figure 3**, **Supplementary Note 4**). All three approaches also provide a rigorous framework for quantifying sources of uncertainty and have already been applied successfully to conventionally acquired datasets (**Box 3**).

Box 3. Current Practice in Community Modelling

Ovaskainen et al.[71] used a joint species distribution model to predict the distributions of 55 butterfly species scored for presence/absence on a grid of 2609 10 X 10-km cells across Great Britain that had been sampled from 1995-1999 in a large citizen-science project. The model was successfully parameterised with a training dataset of just 300 cells and four environmental covariates (degree-days and three types of vegetation cover), plus spatially structured latent variables. Latent variables use observed species subgroupings to detect the effects of unmeasured environmental filters or species interactions such as competition. The parameterised model was used to predict butterfly communities in the testing dataset, which consisted of the remaining 2309 grid cells. Together, the measured and latent variables explained an average of 42% of the variance in species occurrence (with medium-prevalence species more accurately predicted), and the two most dominant latent variables revealed a north-south gradient in species composition, with especially distinct communities in the southeast and northwest. Species richness per grid cell was accurately predicted, and the model's ability to discriminate presence and absence was high (mean AUC = 0.91).

Kéry and Royle[75] used community-occupancy modelling to analyse the 2001 Swiss breeding-bird survey while accounting for variation in detectability due to season, site, and species effects. The dataset consisted of 254 1-km^2 grid cells, each visited three times. The fitted model predicted each species' probability of occurrence as a function of site elevation and forest cover, as well as variance in the uncertainty of occurrence estimates, making it possible to estimate species distributions across the landscape and confidence in those estimates. Parameter estimates were naturally less precise for rare species, but information could be 'borrowed' from data-rich species to increase the precision of predictions for rare species. These procedures were able to compensate for the fact that only 134 total bird species had been detected in the survey, which is less than the true total of 163 species known to breed regularly in Switzerland, plus 22 occasional residents (the testing dataset). The occupancy-corrected model estimated that between 1 and 11 species had been overlooked per grid cell and thus, that the true total in 2001 was 169 species.

Mokany et al.[76] applied Generalised Dissimilarity Modelling (GDM) to a dataset of 2330 expert surveys of New Zealand land snails, which recorded 845 of 998 known species. The GDM was parameterised with a training dataset of 2280 surveys and fourteen environmental variables and explained 57% of the variation in beta diversity. In addition, a generalised additive model parameterised on the training dataset explained 27% of the variation in species richness (after scaling the 20 x 20-m survey quadrats to match the area of modelling units (200 x 200-m); see discussion of scaling in **Supplementary Note 4**). Finally, the outputs were combined using a procedure called DynamicFOAM to assign snail species to communities across New Zealand. Error was assessed by predicting compositions in a testing dataset of 50 sites that had been held out of the model. On average, the model was able to predict half the species that had been observed in each cell, and the predicted total occupancy area per species was highly correlated with the number of quadrat occurrences (Pearson's r = 0.902). When quadrats were pooled into groups of 3 to 400 to reduce sampling stochasticity, predicted species richnesses almost perfectly explained observed richnesses (R^2 = 0.99).

By mapping species distributions as the primary output, we do not lock ourselves into an arbitrary set of convenient indicators, and ongoing discoveries on the relationship between biodiversity and function, which are typically carried out at the species level, can be added. As an illustration, the species diversity of wood-decaying fungi in natural forests is notoriously difficult to assay but can be predicted in part by the volume and species diversity of the stock of dead wood on the ground[77], and these environmental covariates are partially quantifiable via airborne LiDAR sensors (**Supplementary Note 1**)[78], thus allowing EO-based inference of the distribution and level of wood-decaying fungal diversity. Subsequent and unrelated research has suggested that pieces of dead wood inhabited by a higher diversity of fungal species decompose more slowly, possibly due to more intense interference competition[34]. Combining the two results suggests that an EO-derived map of fungal species diversity could be used to contrast landscape management options for how well they conserve saprotrophic fungal biodiversity and thus enhance carbon storage.

Two further reasons for focusing on species-resolution maps as the primary output are that the regional species pool (*gamma diversity*) and the biological dissimilarity of sites (*beta diversity*) could contribute to maintaining functional stability[58,79,80] and that species-resolution outputs retain the option of aggregation to represent different aspects of biodiversity, including higher-taxonomic, functional, and phylogenetic groupings[81].

Many methods are also available to predict *individual* species ranges, and EO can help improve their accuracy, as shown by an example[82] combining MODIS satellite data with environmental DNA to map an invasive diatom over a watershed [Target 9, invasive species pathway identified] (**Supplementary Figure 3.1**). However, ecosystem functions and services are rarely delivered by only one species, and simply summing the outputs of individual models to simulate communities is computationally inefficient, statistically flawed, and does not account for species interactions[83].

From CEOBE to Aichi

In essence, our argument is that new technologies make the new community-modelling approaches (**Box 3, Figure 3**) widely feasible, especially in biodiversity hotspots, where it is particularly difficult to generate large datasets. Larger numbers of environmental covariates and species together increase explanatory power by providing a greater breadth of predictors, and by exploiting latent variables and letting rare species 'borrow' information[42,75,84], respectively. As a result, continuous streams of EO data can be more powerfully interpreted to track biodiversity status and trends (**Figure 1**).

The predictive performance of fitted models can be cross-validated by rounds of comparison with testing datasets that were either split from the model-training dataset[71,76] or derived from historical and expert knowledge[75], and thus, the adequacy of the input data and sampling design, or conversely the degree of model uncertainty, can be assessed *post hoc* (**Box 3**). The regularly updated biodiversity maps that are the primary outputs of the CEOBE approach (**Figure 1**), plus the quantified uncertainty in those maps, can then be incorporated into a larger process of structured decision making and adaptive management[85-87] to (1) identify likely consequences of proposed actions by observing natural experiments that mimic those actions, (2) compare observed results of management interventions against objectives, and (3) help identify and tackle sources of uncertainty.

An early example of the CEOBE approach is given by Sollmann et al.[42], who used community-occupancy modelling to connect environmental covariates from the 5-m-resolution RapidEye satellite to point-sample data from camera traps in three tropical-forest logging concessions in Sabah, Malaysian Borneo, one of which has been managed to reduced-impact-logging standards set by the Forest Stewardship Council (Aichi Target 7, sustainable management under forestry). The dataset consisted of detection events for 28 mammal species at 166 camera-trap stations, each station scored using EO data for distance to water, distance to oil-palm plantation, and forest condition. Estimated relationships between species occurrence and the three covariates were used to predict species occurrence across the three reserves, with rare mammal species borrowing information from more common ones. Species richness was estimated to be higher in the FSC-certified reserve, particularly for threatened species (Target 12, improved conservation status of threatened species). The percentage of area occupied, which could indicate larger population sizes, was also estimated to be higher in the FSC-certified reserve for the majority of species, including for some highly endangered species like the Sunda pangolin *Manis javanica*. Finally, the modelled species richness maps were found to correlate strongly with EO-estimated aboveground biomass at the large spatial grain of whole reserves, but not at a finer resolution (potentially due to hunting at reserve borders), further demonstrating the critical contribution of ground-level point samples for linking pure-EO data to biodiversity.

The major remaining components of uncertainty relate to generalisability, because only a single FSC-certified reserve was sampled; the applicability of results to arboreal species, which tend to be detected more frequently in forests with disturbed canopy but are not necessarily more widespread in these forests; and wide confidence intervals around parameter estimates for some species as a consequence of sparse data and a fairly complex hierarchical model. This example serves as a proof of concept that camera trapping and occupancy modelling can be used to assess biodiversity conservation based on species maps, and the approach has been incorporated in the ten-year forest management plan and wildlife monitoring strategy for the FSC-certified area. Repeated surveys will help to narrow uncertainties in the model, and a future power analysis is planned to estimate the sampling effort required to detect trends and/or provide estimates with a desired level of certainty[88].

Another example of the CEOBE approach is the use of Generalised Dissimilarity Modelling to connect EO-derived metrics of habitat degradation and fragmentation[89,90] to over 300 million records of more than 400,000 species from the Global Biodiversity Information Facility (www.gbif.org) and the Map of Life (mol.org)[91]. The GDM models spatial turnover in biodiversity composition at 1-km-resolution globally, and by invoking the assumption that terrestrial biodiversity declines according to the classical species-area power function, the GDM estimates the proportion of biodiversity that has been retained in each grid cell after habitat loss, based on the proportion of similar habitat remaining unimpacted within the landscape[92]. This metric thus tracks whether rates of loss, degradation, and fragmentation of natural habitats are being reduced (Aichi Target 5). Further, by combining this approach with a global database of protected-area coverage (www.protectedplanet.net), it is possible to report progress against Target 11, which aims for protected areas to cover areas of particular importance to biodiversity and ecosystem services and to be ecologically representative and connected (see also Ref. 93). An important caveat is that the biodiversity data in this case are historical in nature and thus contain the taxonomic and sampling biases and constraints of the past (**Box 2**). Ideally, the biodiversity data will transition to up-to-date, properly sampled, and more taxonomically comprehensive point samples.

Of course, CEOBE outputs cannot contribute to all Aichi Targets, namely those that are focused on policy, planning, and funding reform (Targets 2, 3, 4, 20), the conservation of genetic cultivars (Target 13), the alleviation of climate-change pressures on coral reefs (Target 10), benefits sharing (Target 16), and the integration of traditional knowledge (Target 18). It also remains to be seen how well or poorly EO data reflect biodiversity in aquatic ecosystems (Targets 6 and 11), although environmental DNA on its own is a highly promising source of data on aquatic biodiversity. On the other hand, the efficient production of biodiversity maps and open access to analytical pipelines will help to disseminate the science base and technologies related to biodiversity (Target 19), and could contribute to public awareness of efforts to conserve biodiversity (Target 1) and improve the efficiency of national biodiversity planning (Target 17).

Conclusions

It is extremely difficult to identify all the species present in a location (*the Linnaean challenge*), to delimit the geographic distributions of species (*the Wallacean challenge*), and to quantify their responses to natural and anthropogenic environmental change (*the Hutchinsonian challenge*)[94]. A synergy of Earth Observation, automated recording devices, high-throughput DNA sequencing, and modern statistical modelling can meet these challenges by making it possible to scale up from data-rich but finite sets of point samples to spatially continuous biodiversity maps, which are more informative than a few convenient indicator species but still let us generate summary statistics to communicate trends to decision-makers and the general public. The use of formal statistical frameworks lets us quantify error, identify gaps in our understanding, objectively rank the most likely pressures on biodiversity from multiple candidates, and increase the robustness of change detection. Adding information on species interactions and functions helps link biodiversity to ecosystem functions and services (**Box 1, Figure 1**) in a process-based approach[56], rather than relying on crude estimates from land classes[35]. Finally, as DNA-based technologies mature, the same samples could track population-genetic diversity[64,95,96].

A global, multi-resolution monitoring network is thus within our reach but will still involve a number of challenges associated with technical capacity, computation and data storage, and data standardisation. For every ecologically distinct region, there will be an initial cost to collect data for model parameterisation, followed by a low level of continuous sampling, which will be necessary for updating models and for surveillance monitoring of environmental drivers that are invisible to EO, such as broad-spectrum insecticides. The initial costs are probably best borne by governments, as part of their commitment to the Convention on Biological Diversity, and there is great promise in using citizen-science networks to collect standardised, bulk biodiversity samples over large areas. A laudable example is the School Malaise Trap Program that recruited hundreds of secondary-school science classes to collect arthropods across Canada (malaiseprogram.com). Initial investment could also come from existing monitoring budgets with the expectation that additional information content will compensate for reduced sample numbers within existing programs[82]. The follow-up continuous sampling requires steady funding streams, and the standardisation of the CEOBE approach meets the needs of international certification schemes, such as REDD+, Climate, Community & Biodiversity Standards, Forest Stewardship Council, and the Roundtable on Sustainable Palm Oil, which all require the continuous monitoring of biodiversity and ecosystem services. Biodiversity-offset payments to mitigate the impacts of development and carbon emissions are also expected to provide funding streams, and standardised assessments are needed to ensure that offsetting results in biodiversity net gain[97].

The CEOBE approach also depends on institutional support for the multidisciplinary collaborations needed to generate, combine, analyse, and act upon data from disparate disciplines (EO, ARDs, genomics, taxonomy and systematics, ecosystem functions and services, statistics, and decision science), expertise that no single individual has[12,30,98]. Identifying causal determinants of species distributions needs a clear understanding of phylogenetic structure and functional diversity, the ecological processes involved, and what EO sensors can and cannot observe[99]. Expert knowledge will also contribute to sampling design and covariate selection so that the full breadth of environmental conditions is captured, especially those not visible to EO.

On the other hand, collaborations need not be global. Political and social interests will vary by region, and agencies should be encouraged to trial CEOBE within their jurisdictions where there are clear opportunities to improve management, while also enforcing the publication of primary data and analytical pipelines[27,100]. The Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) could play an important role as a global coordinating institution.

Resources for environmental management are always likely to be limited, but by doing more with our expensively gained field data, we can take action more efficiently and effectively. What is required now is leadership by governments and international organisations to stimulate integrated research and to endorse the use of comprehensive biodiversity information[6].

Acknowledgements

This article is a product of the EO-BESS Working Group, organised by Heiko Balzter, David Raffaelli, and Beth Cole and funded by the UK Natural Environment Research Council. Individual author acknowledgements are in Supplementary Information.

Author Contributions

BC and HB led the sections on Earth Observation technology. KB and DWY led the sections on Biodiversity technology. AB led the sections on Statistical modelling. AB, RS, AW, OO, and DWY led the sections on case studies (Box 3 and CEOBE to Aichi). CM led the Conclusions section. Figures were created by KB, AB, CC, and AZ. All authors contributed to multiple rewrites, with a large contribution by DR. AB and DWY wrote the first draft and supervised the work.

Additional Information

Correspondence should be addressed to DWY.

Competing Interests

DWY and AV are co-founders of a private company that provides commercial metabarcoding services.

References

1 Verrelst, J. *et al.* Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing* **108**, 273-290, doi:10.1016/j.isprsjprs.2015.05.005 (2015).

2 Wulder, M. A. *et al.* Virtual constellations for global terrestrial monitoring. *Remote Sensing of Environment* **170**, 62-76, doi:10.1016/j.rse.2015.09.001 (2015).

3 Toth, C. & Jóźków, G. Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing* **115**, 22-36, doi:10.1016/j.isprsjprs.2015.10.004 (2016).

4 O'Connor, B. *et al.* Earth observation as a tool for tracking progress towards the Aichi Biodiversity Targets. *Remote Sensing in Ecology and Conservation* **1**, 19-28, doi:10.1002/rse2.4 (2015).

5 Skidmore, A. K. *et al.* Environmental science: Agree on biodiversity metrics to track from space. *Nature News* **523**, 403, doi:10.1038/523403a (2015).

6 **Pettorelli, N. *et al.* Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sensing in Ecology and Conservation* 2, 122-131, doi:10.1002/rse2.15 (2016). Identifies candidate essential biodiversity variables derived directly from remote sensing and assesses their feasibility for global biomontoring.**

7 CBD. Decision adopted by the Conference of the Parties to the Convention on Biological Diversity at its Tenth Meeting. Decision X/2. The Strategic Plan for Biodiversity 2011–2020 and the Aichi Biodiversity Targets. (UNEP/CBD/COP/DEC/X/2. 29 October 2010., 2010).

8 UNGA. Resolution adopted by the General Assembly on 25 September 2015. Transforming our world: the 2030 Agenda for Sustainable Development. United Nations General Assembly. Seventieth Session. A/RES/70/1., (2015).

9 UNFCCC. Adoption of the Paris Agreement. FCCC/CP/2015/L.9/Rev.1. (2015).

10 Tittensor, D. P. *et al.* A mid-term analysis of progress toward international biodiversity targets. *Science* **346**, 241-244, doi:10.1126/science.1257484 (2014).

11 Durance, I. *et al.* The Challenges of Linking Ecosystem Services to Biodiversity. *Advances in Ecological Research* **54**, 87-134, doi:10.1016/bs.aecr.2015.10.003 (2016).

12 Pettorelli, N., Owen, H. & Duncan, C. How do we want Satellite Remote Sensing to support biodiversity conservation globally? *Methods in Ecology and Evolution* **7**, 656-665, doi:10.1111/2041-210X.12545 (2016).

13 IPBES. Decision and scoping report for the IPBES global assessment on biodiversity and ecosystem services. Decision IPBES-4/1. Work Programme of the Platform. (2016).

14 Dawson, T. P., Cutler, M. E. J. & Brown, C. The role of remote sensing in the development of SMART indicators for ecosystem services assessment. *Biodiversity* **17**, 136-148, doi:10.1080/14888386.2016.1246384 (2016).

15 Pereira, H. M. *et al.* Essential Biodiversity Variables. *Science* **339**, 277-278, doi:10.1126/science.1229931 (2013).

16 Proença, V. *et al.* Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. *Biological Conservation*, doi:10.1016/j.biocon.2016.07.014 (2016).

17 Peres, C. A., Emilio, T., Schietti, J., Desmoulière, S. J. M. & Levi, T. Dispersal limitation induces long-term biomass collapse in overhunted Amazonian forests. *Proceedings of the National Academy of Sciences USA* **113**, 892–897, doi:10.1073/pnas.1516525113 (2016).

18 Levi, T., Shepard Jr, G. H., Ohl-Schacherer, J., Peres, C. A. & Yu, D. W. Modelling the long-term sustainability of indigenous hunting in Manu National Park, Peru: landscape-scale management implications for Amazonia. *Journal of Applied Ecology* **46**, 804-814, doi:10.1111/j.1365-2664.2009.01661.x (2009).

19 Ferrier, S. Extracting More Value from Biodiversity Change Observations through Integrated Modeling. *BioScience* **61**, 96-97, doi:10.1525/bio.2011.61.2.2 (2011).

20 Belward, A. S. & Skøien, J. O. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing* **103**, 115-128, doi:10.1016/j.isprsjprs.2014.03.009 (2015).

21 Roy, D. P. *et al.* Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment* **145**, 154-172, doi:10.1016/j.rse.2014.02.001 (2014).

22 **Turner, W. *et al.* Free and open-access satellite data are key to biodiversity conservation. *Biological Conservation* 182, 173-176, doi:10.1016/j.biocon.2014.11.048 (2015).**

Identification of continuity, affordability, and accessibility as having major

impact on the utility of remote sensing data for tracking and understanding biodiversity change.

23 Butler, D. Earth observation enters next phase. *Nature* **508**, 160-161, doi:10.1038/508160a (2014).

24 Berger, M., Moreno, J., Johannessen, J. A., Levelt, P. F. & Hanssen, R. F. ESA's sentinel missions in support of Earth system science. *Remote Sensing of Environment* **120**, 84-90, doi:10.1016/j.rse.2011.07.023 (2012).

25 Malenovský, Z. *et al.* Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment* **120**, 91-101, doi:10.1016/j.rse.2011.09.026 (2012).

26 **Asner, G. P. *et al.* Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science* 355, 385-389, doi:10.1126/science.aaj1987 (2017).**
Large-scale mapping of multivariate forest canopy functional traits from airborne laser-guided imaging spectroscopy.

27 Petrou, Z. I., Manakos, I. & Stathaki, T. Remote sensing for biodiversity monitoring: a review of methods for biodiversity indicator extraction and assessment of progress towards international targets. *Biodiversity and Conservation* **24**, 2333-2363, doi:10.1007/s10531-015-0947-z (2015).

28 Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R. & Woodcock, C. E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment* **122**, 2-10, doi:10.1016/j.rse.2012.01.010 (2012).

29 Lindenmayer, D. B. & Likens, G. E. Direct Measurement Versus Surrogate Indicator Species for Evaluating Environmental Change and Biodiversity Loss. *Ecosystems* **14**, 47-59, doi:10.1007/s10021-010-9394-6 (2011).

30 Mueller, M. & Geist, J. Conceptual guidelines for the implementation of the ecosystem approach in biodiversity monitoring. *Ecosphere* **7**, e01305, doi:10.1002/ecs2.1305 (2016).

31 Newton, A. C. Implications of Goodhart's Law for monitoring global biodiversity loss. *Conservation Letters* **4**, 264-268, doi:10.1111/j.1755-263X.2011.00167.x (2011).

32 Smaldino, P. E. & McElreath, R. The natural selection of bad science. *Royal Society Open Science* **3**, doi:10.1098/rsos.160384 (2016).

33 Crowther, T. W. *et al.* Biotic interactions mediate soil microbial feedbacks to climate change. *Proceedings of the National Academy of Sciences* **112**, 7033-7038, doi:10.1073/pnas.1502956112 (2015).

34 Yang, C. Y. *et al.* Higher fungal diversity is correlated with lower CO_2 emissions from dead wood in a natural forest. *Scientific Reports* **6**, doi:10.1038/srep31066 (2016).

35 Eigenbrod, F. *et al.* The impact of proxy-based methods on mapping the distribution of ecosystem services. *Journal of Applied Ecology* **47**, 377-385, doi:10.1111/j.1365-2664.2010.01777.x (2010).

36 Snaddon, J., Petrokofsky, G., Jepson, P. & Willis, K. J. Biodiversity technologies: tools as change agents. *Biology Letters* **9**, 20121029, doi:10.1098/rsbl.2012.1029 (2013).

37 Turner, W. Sensing biodiversity. *Science* **346**, 301-302, doi:10.1126/science.1256014 (2014).

38 Acevedo, M. A. & Villanueva-Rivera, L. J. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* **34**, 211-214, doi:10.2193/0091-7648(2006)34[211:UADRSA]2.0.CO;2 (2006).

39 Lammers, M. O., Brainard, R. E., Au, W. W. L., Mooney, T. A. & Wong, K. B. An ecological acoustic recorder (EAR) for long-term monitoring of biological and anthropogenic sounds on coral reefs and other marine habitats. *The Journal of the Acoustical Society of America* **123**, 1720-1728, doi:10.1121/1.2836780 (2008).

40 Jung, K. & Kalko, E. K. V. Adaptability and vulnerability of high flying Neotropical aerial insectivorous bats to urbanization. *Diversity and Distributions* **17**, 262-274, doi:10.1111/j.1472-4642.2010.00738.x (2011).

41 Aide, T. M. *et al.* Real-time bioacoustics monitoring and automated species identification. *PeerJ* **1**, e103, doi:10.7717/peerj.103 (2013).

42 Sollmann, R. *et al.* Quantifying mammal biodiversity co-benefits in certified tropical forests. *Diversity and Distributions* **23**, 317-328, doi:10.1111/ddi.12530 (2017).

43 Yu, D. W. *et al.* Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* **3**, 613–623, doi:10.1111/j.2041-210X.2012.00198.x (2012).

44 Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. Environmental DNA. *Molecular Ecology* **21**, 1789-1793, doi:10.1111/j.1365-294X.2012.05542.x (2012).

45 Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution* **29**, 358-367, doi:10.1016/j.tree.2014.04.003 (2014).

46 **Ji, Y. *et al.* Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* 16, 1245-1257, doi:10.1111/ele.12162 (2013).**

Demonstrated that metabarcoding data can be as reliable as high-quality morphological datasets for environmental management decisions.

47 Lejzerowicz, F. *et al.* High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Scientific Reports* **5**, 13932, doi:10.1038/srep13932 (2015).

48 Edwards, D. P. *et al.* Selective-logging and oil palm: multitaxon impacts, biodiversity indicators, and trade-offs for conservation planning. *Ecological Applications* **24**, 2029-2049, doi:10.1890/14-0010.1 (2014).

49 Chariton, A. A. *et al.* Emergent technologies and analytical approaches for understanding the effects of multiple stressors in aquatic environments. *Marine and Freshwater Research* (2015).

50 Aylagas, E., Borja, Á., Irigoien, X. & Rodríguez-Ezpeleta, N. Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science* **3**, doi:10.3389/fmars.2016.00096 (2016).

51 Visco, J. A. *et al.* Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. *Environmental Science & Technology* **49**, 7597-7605, doi:10.1021/es506158m (2015).

52 Xue, K. *et al.* Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming. *Nature Clim. Change* **6**, 595-600, doi:10.1038/nclimate2940 (2016).

53 Asner, G. P., Knapp, D. E., Anderson, C. B., Martin, R. E. & Vaughn, N. Large-scale climatic and geophysical controls on the leaf economics spectrum. *Proceedings of the National Academy of Sciences* **113**, E4043–E4051, doi:10.1073/pnas.1604863113 (2016).

54 Fisher, J. B., Sweeney, S. & Brzostek, E. R. Tree–mycorrhizal associations detected remotely from canopy spectral properties. *Global Change Biology* **22**, 2596-2607, doi:10.1111/gcb.13264 (2016).

55 **Bohan, D. A. *et al.* Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends in Ecology & Evolution*, doi:10.1016/j.tree.2017.03.001 (2017).**
Proposal to infer interaction networks and species functions from co-occurrence data generated by high-throughput biodiversity methods.

56 Barnes, A. D. *et al.* Species richness and biomass explain spatial turnover in ecosystem functioning across tropical and temperate ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **37**, doi:10.1098/rstb.2015.0279 (2016).

57 Brose, U. & Hillebrand, H. Biodiversity and ecosystem functioning in dynamic landscapes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150267, doi:10.1098/rstb.2015.0267 (2016).

58 Burley, H. M., Mokany, K., Ferrier, S. & Laffan, S. W. Macroecological scale effects of biodiversity on ecosystem functions under environmental change. *Ecol. Evol.* **6**, 2579-2593, doi:10.1002/ece3.2036 (2016).

59 Tang, M. *et al.* High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution* **6**, 1034-1043, doi:10.1111/2041-210X.12416 (2015).

60 Wood, T. J., Holland, J. M. & Goulson, D. Providing foraging resources for solitary bees on farmland: current schemes for pollinators benefit a limited suite of species. *Journal of Applied Ecology* **54**, 323-333, doi:10.1111/1365-2664.12718 (2017).

61 McIntyre, P. B., Jones, L. E., Flecker, A. S. & Vanni, M. J. Fish extinctions alter nutrient recycling in tropical freshwaters. *Proceedings of the National Academy of Sciences* **104**, 4461-4466, doi:10.1073/pnas.0608148104 (2007).

62 Solan, M. *et al.* Extinction and Ecosystem Function in the Marine Benthos. *Science* **306**, 1177-1180, doi:10.1126/science.1103960 (2004).

63 Sunarto, Sollmann, R., A., M. & Kelly, M. J. Camera trapping for the study and conservation of tropical carnivores. *Raffles Bulletin of Zoology* **28**, 21–42 (2013).

64 Sigsgaard, E. E. *et al.* Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution* **1**, 0004, doi:10.1038/s41559-016-0004 (2016).

65 Ferrier, S. Mapping Spatial Pattern in Biodiversity for Regional Conservation Planning: Where to from Here? *Systematic Biology* **51**, 331-363 (2002).

66 Ferrier, S. & Guisan, A. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* **43**, 393-404, doi:10.1111/j.1365-2664.2006.01149.x (2006).

67 Honrado, J. P., Pereira, H. M. & Guisan, A. Fostering integration between biodiversity monitoring and modelling. *Journal of Applied Ecology* **53**, 1299-1304, doi:10.1111/1365-2664.12777 (2016).

68 D'Amen, M., Rahbek, C., Zimmermann, N. E. & Guisan, A. Spatial predictions at the community level: from current approaches to future frameworks. *Biological Reviews* **92**, 169-187, doi:10.1111/brv.12222 (2017).

69 Warton, D. I. *et al.* So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution* **30**, 766-779, doi:10.1016/j.tree.2015.09.007 (2015).

70 Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution* **7**, 549-555, doi:10.1111/2041-210X.12501 (2016).

71 Ovaskainen, O., Roy, D. B., Fox, R. & Anderson, B. J. Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution* **7**, 428-436, doi:10.1111/2041-210X.12502 (2016).

72 **Ovaskainen, O. *et al.* How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, doi:doi:10.1111/ele.12757 (2017).**
Presents a general joint species distribution modelling framework for examining how the link from environmental covariates to species occurrence depends on species traits and phylogenetic relationships.

73 Dorazio, R. M. & Royle, J. A. Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association* **100**, 389-398 (2005).

74 Ferrier, S., Manion, G., Elith, J. & Richardson, K. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions* **13**, 252-264, doi:10.1111/j.1472-4642.2007.00341.x (2007).

75 Kery, M. & Royle, A. J. in *Modeling demographic processes in marked populations. Environmental and ecological statistics, Vol. 3.* (eds D. L. Thomson, E. G. Cooch, & M. J. Conroy) (Springer, 2009).

76 Mokany, K., Harwood, T., Overton, J., Barker, G. & Ferrier, S. Combining α - and β - diversity models to fill gaps in our knowledge of biodiversity. *Ecology letters* **14**, 1043-1051, doi:10.1111/j.1461-0248.2011.01675.x (2011).

77 Hottola, J., Ovaskainen, O. & Hanski, I. A unified measure of the number, volume and diversity of dead trees and the response of fungal communities. *Journal of Ecology* **97**, 1320-1328, doi:10.1111/j.1365-2745.2009.01583.x (2009).

78 Mücke, W., Deák, B., Schroiff, A., Hollaus, M. & Pfeifer, N. Detection of fallen trees in forested areas using small footprint airborne laser scanning data. *Canadian Journal of Remote Sensing* **39**, S32-S40, doi:10.5589/m13-013 (2013).

79 Pasari, J. R., Levi, T., Zavaleta, E. S. & Tilman, D. Several scales of biodiversity affect ecosystem multifunctionality. *Proceedings of the National Academy of Sciences* **110**, 10219-10222, doi:10.1073/pnas.1220333110 (2013).

80 Wang, S. & Loreau, M. Ecosystem stability in space: α , β and γ variability. *Ecology letters* **17**, 891-901, doi:10.1111/ele.12292 (2014).

81 Cardinale, B. J., Duffy, J. E., Gonzalez, A. & Hooper, D. U. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67, doi:10.1038/nature11148 (2012).

82 Olson, J. R., Hawkins, C. P., Mock, K., Huntington, J. & Susfalk., R. System for Mapping And Predicting Species Of Concern (SMAP-SOC), Phase I Final Report and Phase II Plan., (NASA Earth Science Division/Applied Sciences Program, Washington D.C., 2014).

83 Calabrese, J. M., Certain, G., Kraan, C. & Dormann, C. F. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography* **23**, 99-112, doi:10.1111/geb.12102 (2014).

84 Ovaskainen, O. & Soininen, J. Making more out of sparse data: hierarchical modeling of species communities. *Ecology* **92**, 289-295, doi:10.1890/10-1251.1 (2011).

85 Ferretti, V. & Pomarico, S. Ecological land suitability analysis through spatial indicators: An application of the Analytic Network Process technique and Ordered Weighted Average approach. *Ecological Indicators* **34**, 507-519, doi:10.1016/j.ecolind.2013.06.005 (2013).

86 Marcot, B. G. *et al.* Recent advances in applying decision science to managing national forests. *Forest Ecology and Management* **285**, 123-132, doi:10.1016/j.foreco.2012.08.024 (2012).

87 Gregory, R., Long, G., Colligan, M., Geiger, J. G. & Laser, M. When experts disagree (and better science won't help much): Using structured deliberations to support endangered species recovery planning. *Journal of Environmental Management* **105**, 30-43, doi:10.1016/j.jenvman.2012.03.001 (2012).

88 Steidl, R. J., Hayes, J. P. & Schauber, E. Statistical power analysis in wildlife research. *Journal of Wildlife Management* **61**, 270-279 (1997).

89 Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850-853, doi:10.1126/science.1244693 (2013).

90 Newbold, T. *et al.* Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science* **353**, 288-291, doi:10.1126/science.aaf2201 (2016).

91 GEO BON. (ed Group on Earth Observations Biodiversity Observation Network Secretariat) (Leipzig, 2016).

92 Allnutt, T. F. *et al.* A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. *Conservation Letters* **1**, 173-181, doi:10.1111/j.1755-263X.2008.00027.x (2008).

93 Ferrier, S. *et al.* Mapping more of terrestrial biodiversity for global conservation assessment. *BioScience* **54**, 1101-1109 (2004).

94 Cardoso, P., Erwin, T. L., Borges, P. A. V. & New, T. R. The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation* **144**, 2647-2655 (2011).

95 Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters* **18**, 1-16, doi:10.1111/ele.12376 (2015).

96 Crampton-Platt, A., Yu, D. W., Zhou, X. & Vogler, A. P. Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience* **5**, 1-11, doi:10.1186/s13742-016-0120-y (2016).

97 Maron, M., Gordon, A., Mackey, B., Posssingham, H. P. & Watson, J. E. M. Stop misuse of biodiversity offsets. *Nature* **523**, 401–403 (2015).

98 Palumbo, I. *et al.* Building capacity in remote sensing for conservation: present and future challenges. *Remote Sensing in Ecology and Conservation* **3**, *21-29*, doi:10.1002/rse2.31 (2016).

99 Dafforn, K. A., Johnston, E. L. & Ferguson, A. Big data opportunities and challenges for assessing multiple stressors across scales in aquatic ecosystems. *Marine and Freshwater Research* **67**, 393-413 (2015).

100 Schmeller, D. S. *et al.* Towards a global terrestrial species monitoring program. *Journal for Nature Conservation* **25**, 51-57, doi:10.1016/j.jnc.2015.03.003 (2015).

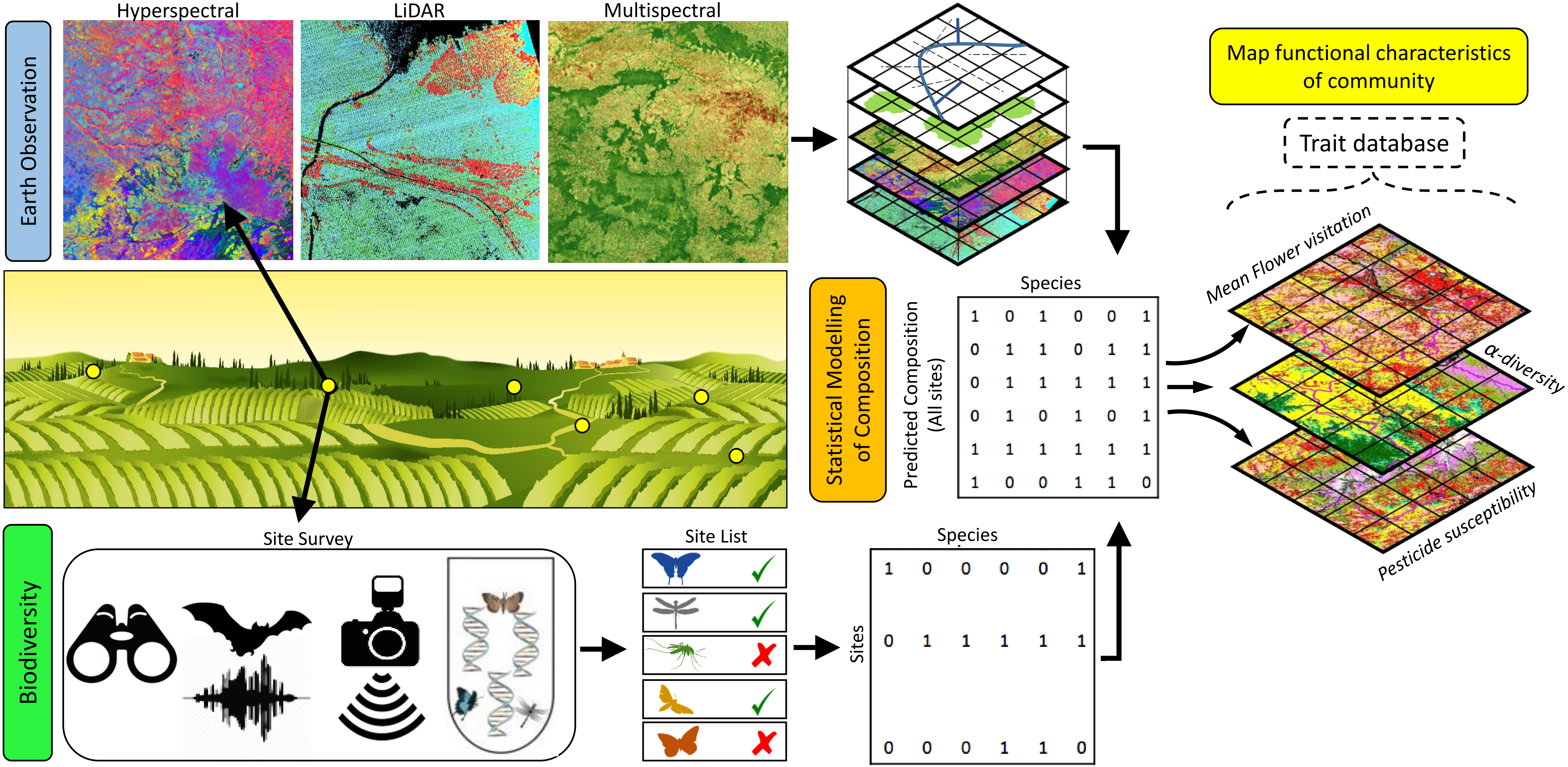
Figure legends

Figure 1. CEOBE – Connecting Earth Observation to Biodiversity and Ecosystems. Top row left: EO data and other geographical datasets are used to generate spatially continuous maps of biophysical data (**S1, S2**). **Middle row left**: A real landscape with point-sample locations indicated by yellow dots. **Bottom row left**: Biodiversity is recorded manually using traditional methods, automated audio or image recording devices, or metabarcoding or metagenomic pipelines to generate a site X species table (**Figure 2**, **S3**). However, most of the landscape is not sampled (empty rows in the table). **Right side**: The point samples are combined statistically with continuous biophysical maps to predict biodiversity composition over the whole landscape (**S4**). In combination with ancillary data like trait databases, process-based models can then identify the functional composition of any location and map the expected distributions of ecosystem functions and services.

Figure 2. Metabarcoding and metagenomic processing pipelines for high-throughput biodiversity surveys. Top row: Point locations across a landscape are sampled for biodiversity, and DNA is separately extracted from each sample. Three common sample types are (i) bulk samples of arthropods (depicted here), (ii) environmental DNA (eDNA)

from soil, water, and air, and (iii) invertebrate collectors of vertebrate DNA (iDNA), such as mosquitoes, leeches, flies, dung beetles, and ticks. **Left column**: Metabarcoding – Each sample's DNA is amplified via PCR (polymerase chain reaction) for a particular marker gene that is taxonomically informative, the samples are pooled and sequenced on a high-throughput sequencer, and then sorted back to sample by the sample-specific tags added during PCR. The sequences are then clustered into Operational Taxonomic Units (OTUs), which are species hypotheses, and assigned taxonomies by matching against online databases. **Right column**: Meta/mitogenomics – Each sample's total DNA is sequenced, and the output DNA reads are matched to reference genomes, which are often mitochondrial genomes. **Bottom row**: The output of both processing pipelines is a 'sample X species' table. Metabarcoding pipelines are useful for general biodiversity discovery and surveys because online barcode databases are more taxonomically complete, and even without taxonomic assignment, it is possible to calculate community metrics from OTUs only. Metagenomic pipelines are more costly, but advantageous when it is important to reliably identify particular sets of species and to a greater extent preserve relative biomass information. See **S3** for further details. Clip-art courtesy of the Integration and Application Network, University of Maryland Center for Environmental Science (ian.umces.edu/symbols/).

Figure 3. Three statistical pathways to map community composition and summary metrics from the combination of biodiversity point samples and continuous Earth Observation (EO) maps. Local diversity – α; species turnover – β; and regional diversity – γ. For clarity, the figure only considers models for species occurrence (OCC), not abundance. GAM: Generalised Additive Model. DynamicFOAM is described in Ref. 76. See **S4** for further details.



Earth Observation

Hyperspectral

LiDAR

Multispectral

Map functional characteristics of community

Trait database

Statistical Modelling of Composition

Predicted Composition (All sites)

Species

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

Biodiversity

Site Survey

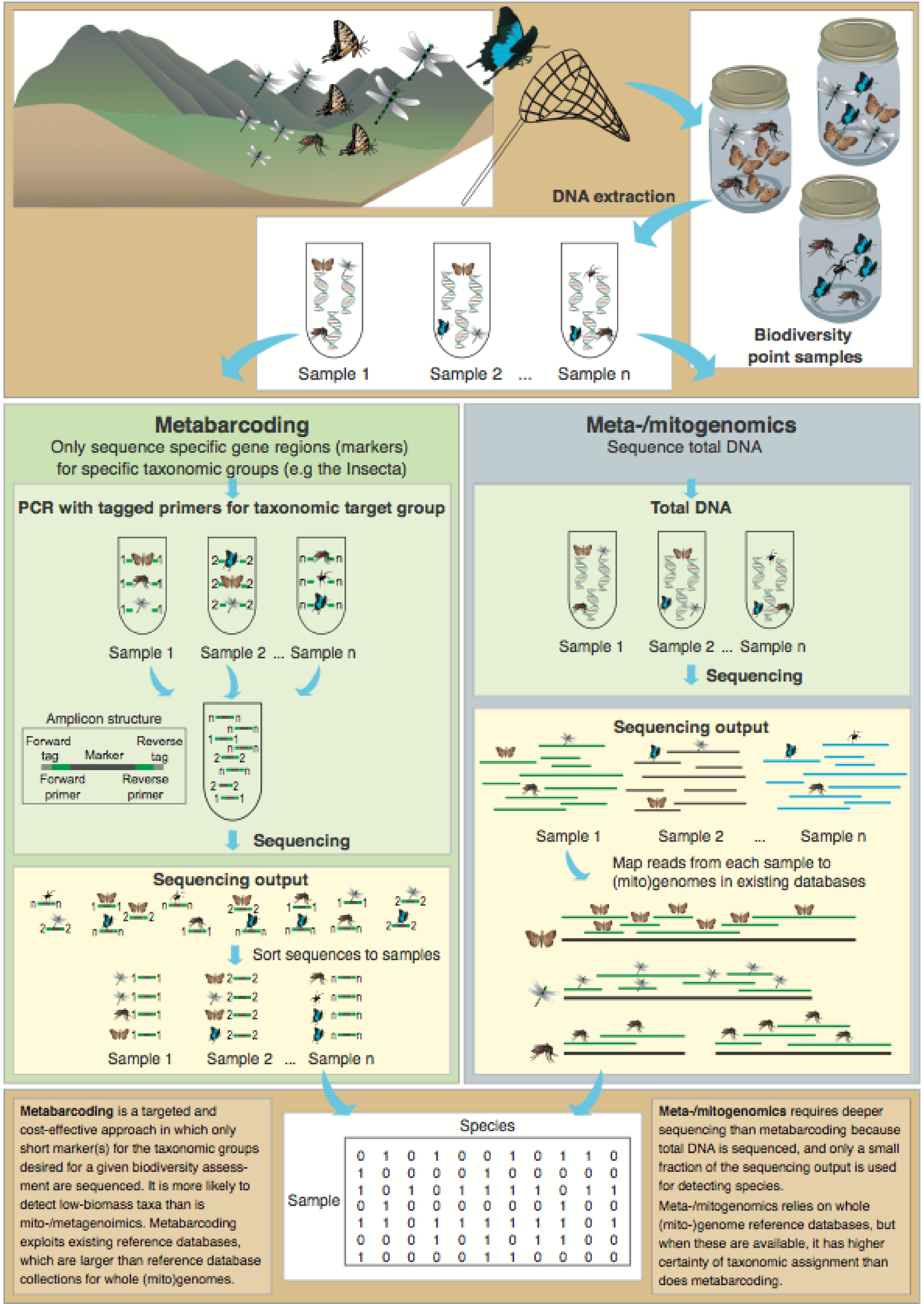
Site List

	✓
	✓
	✗
	✓
	✗

Sites

Species

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

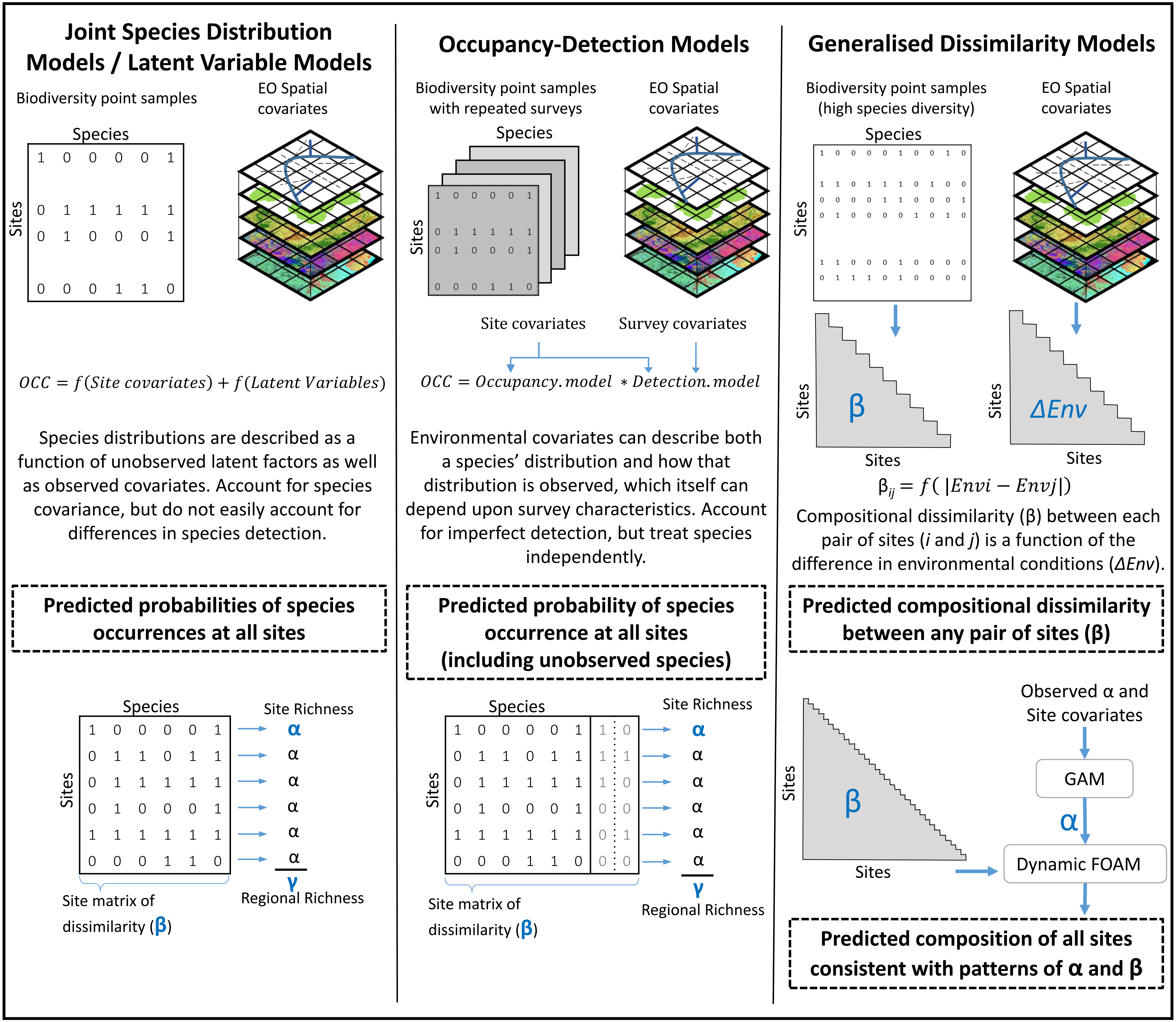


Metabarcoding is a targeted and cost-effective approach in which only short marker(s) for the taxonomic groups desired for a given biodiversity assessment are sequenced. It is more likely to detect low-biomass taxa than is mito-/metagenoimics. Metabarcoding exploits existing reference databases, which are larger than reference database collections for whole (mito)genomes.

	Species										
Sample	0	1	0	1	0	0	1	0	1	1	0
	1	0	0	0	0	1	0	0	0	0	0
	1	0	1	0	1	1	1	1	0	1	1
	0	1	0	0	0	0	0	0	1	0	0
	1	1	0	1	1	1	1	1	1	0	1
	0	0	0	1	0	1	0	1	1	0	0
	1	0	0	0	0	1	1	0	0	0	0

Meta-/mitogenomics requires deeper sequencing than metabarcoding because total DNA is sequenced, and only a small fraction of the sequencing output is used for detecting species.

Meta-/mitogenomics relies on whole (mito-)genome reference databases, but when these are available, it has higher certainty of taxonomic assignment than does metabarcoding.



Joint Species Distribution Models / Latent Variable Models

1 0 0 0 0 1

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

0 0 0 1 1 0

$$OCC = f(Site\ covariates) + f(Latent\ Variables)$$

Species distributions are described as a function of unobserved latent factors as well as observed covariates. Account for species covariance, but do not easily account for differences in species detection.

Predicted probabilities of species occurrences at all sites

1 0 0 0 0 1 → α
0 1 1 0 1 1 → α
0 1 1 1 1 1 → α
0 1 0 0 0 1 → α
1 1 1 1 1 1 → α
0 0 0 1 1 0 → α

γ

Occupancy-Detection Models

1 0 0 0 0 1

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

0 0 0 1 1 0

$$OCC = Occupancy.model * Detection.model$$

Environmental covariates can describe both a species' distribution and how that distribution is observed, which itself can depend upon survey characteristics. Account for imperfect detection, but treat species independently.

Predicted probability of species occurrence at all sites (including unobserved species)

1 0 0 0 0 1 1 0 → α
0 1 1 0 1 1 1 1 → α
0 1 1 1 1 1 1 0 → α
0 1 0 0 0 1 0 0 → α
1 1 1 1 1 1 0 1 → α
0 0 0 1 1 0 0 0 → α

γ

Generalised Dissimilarity Models

1 0 0 0 0 1 0 0 1 0

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

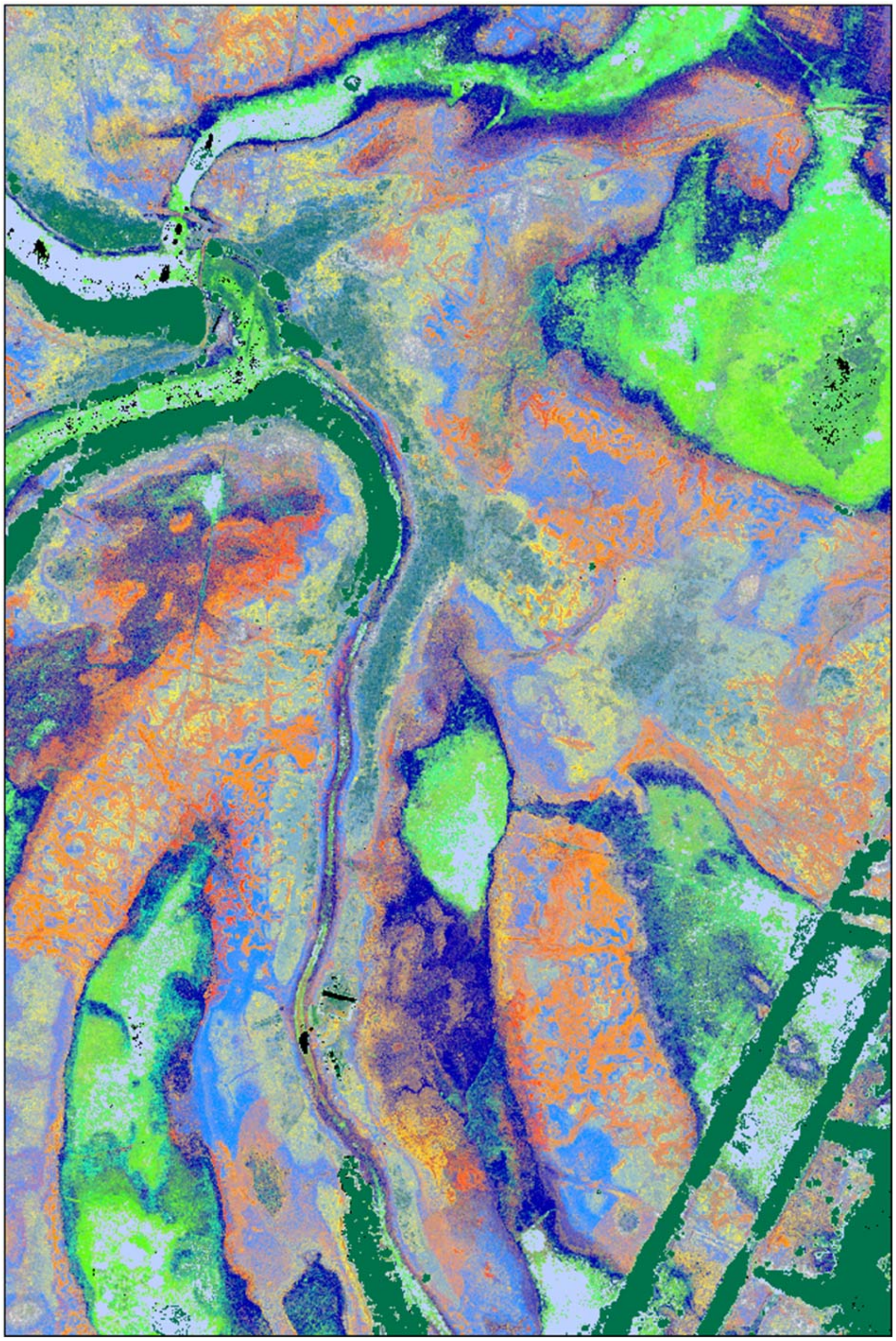
1 1 0 0 0 1 0 0 0 0
0 1 1 1 0 1 0 0 0 0

$$\beta_{ij} = f(\,|Envi - Envj|)$$

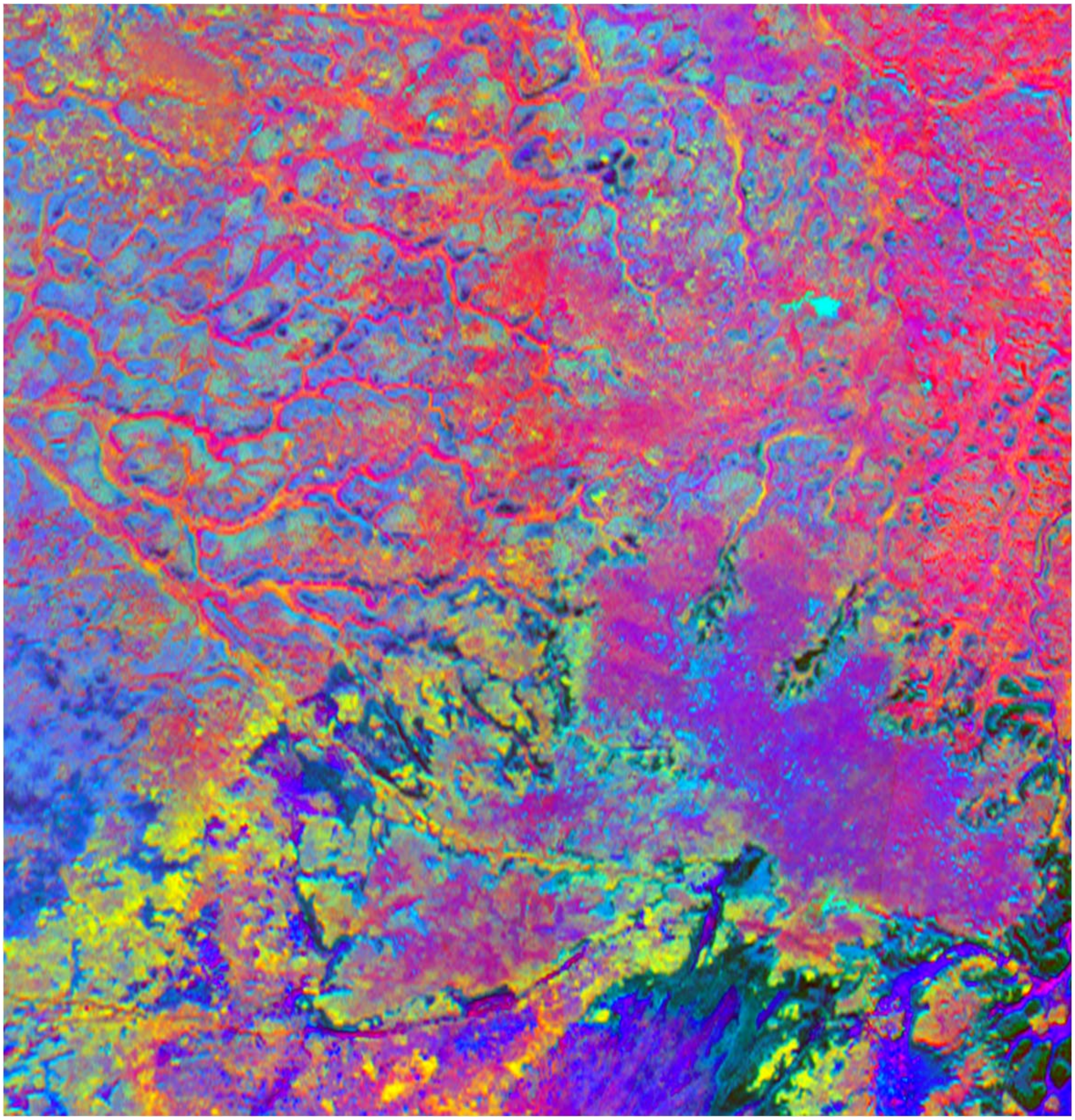
Compositional dissimilarity (β) between each pair of sites (*i* and *j*) is a function of the difference in environmental conditions (*ΔEnv*).

Predicted compositional dissimilarity between any pair of sites (β)

Predicted composition of all sites consistent with patterns of α and β



Fuzzy classification of grassland vegetation in an alkaline grassland in Püspökladány, Hungary, based on airborne LIDAR. Colours represent the weighted probability for a given vegetation class in each cell (0.5m2) (photo credit: András Zlinszky).



Vegetation composition of a peatland using Partial Least Square Regression models on a hyperspectral image. The image is a false colour composite showing the predicted abundance of Graminoids (Red), Shrubs (Green), and Bryophytes (Blue) (photo credit: Beth Cole).

A forest elephant “scanned” during a terrestrial laser-based measurement of a tropical rainforest in Gabon 2013 (photo credit: Kim Calders).

Connecting Earth Observation to High-Throughput Biodiversity Data: Supplementary Information

Bush, Alex[1,2,3]; Sollmann, Rahel[4]; Wilting, Andreas[5]; Bohmann, Kristine[6,7]; Cole, Beth[8]; Balzter, Heiko[8,9]; Martius, Christopher[10]; Zlinszky, András[11]; Calvignac-Spencer, Sébastien[12]; Cobbold, Christina A.[13]; Dawson, Terence P.[14]; Emerson, Brent C.[15,7]; Ferrier, Simon[3]; Gilbert, M. Thomas P.[6,16]; Herold, Martin[17]; Jones, Laurence[18]; Leendertz, Fabian H.[12]; Matthews, Louise[13]; Millington, James D.A.[14]; Olson, John R.[19]; Ovaskainen, Otso[20,21]; Raffaelli, Dave[22]; Reeve, Richard[13]; Rödel, Mark-Oliver[23]; Rodgers, Torrey W.[24]; Snape, Stewart[25]; Visseren-Hamakers, Ingrid[26]; Vogler, Alfried P.[27,28]; White, Piran C.L.[22]; Wooster, Martin J.[14]; Yu, Douglas W.[1,7, ǂ]

1. State Key Laboratory of Genetic Resources and Evolution, **Kunming Institute of Zoology**, Chinese Academy of Sciences, Kunming, Yunnan 650223 China
2. Environment Canada @ Canadian Rivers Institute, Department of Biology, **University of New Brunswick**, PO Box 4400, Fredericton, NB, E3B 5A3, Canada
3. **CSIRO Land and Water**, Canberra, ACT 2601, Australia
4. Department of Wildlife, Fish, & Conservation Biology, 1088 Academic Surge, One Shields Ave, **University of California Davis**, Davis, CA 95616 USA
5. **Leibniz Institute for Zoo and Wildlife Research**, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany
6. EvoGenomics, Natural History Museum of Denmark, **University of Copenhagen**, 1350 Copenhagen K Denmark
7. School of Biological Sciences, **University of East Anglia**, Norwich Research Park, Norwich, Norfolk NR47TJ UK
8. Centre for Landscape and Climate Research and Leicester Institute for Space and Earth Observation (LISEO), **University of Leicester**, University Road, Leicester LE17RH UK
9. NERC National Centre for Earth Observation (NCEO) at **University of Leicester**, University Road, Leicester LE17RH, UK

10. **Center for International Forestry Research (CIFOR)**, P.O. Box 0113 BOCBD, Bogor 16000, Indonesia
11. Balaton Limnological Institute, **Centre for Ecological Research**, Hungarian Academy of Sciences, Tihany 8237 Hungary
12. **Robert Koch Institut**, Berlin 13353 Germany
13. Boyd Orr Centre for Population and Ecosystem Health, **University of Glasgow**, Glasgow G128QQ UK
14. Department of Geography, **King's College London**, Strand Campus, London WC2R2LS UK
15. **IPNA-CSIC**, La Laguna, Tenerife, Canary Islands 38206 Spain
16. NTNU University Museum, **Norwegian University of Science and Technology**, Trondheim 7491 Norway
17. Laboratory of Geo-Information Science and Remote Sensing, **Wageningen University & Research**, Wageningen, the Netherlands
18. Centre for Ecology and Hydrology, **Environment Centre Wales**, Deiniol Road, Bangor LL572UW UK
19. Division of Hydrologic Sciences, **Desert Research Institute**, Las Vegas, NV, USA
20. Department of Biosciences, **University of Helsinki**, Helsinki FI-00014 Finland
21. Centre for Biodiversity Dynamics, Department of Biology, **Norwegian University of Science and Technology**, Trondheim 7491 Norway
22. Environment Department, **University of York**, York YO105NG UK
23. Museum für Naturkunde - **Leibniz Institute for Evolution and Biodiversity Science**, Berlin 10115 Germany
24. Department of Wildland Resources, **Utah State University**, Logan, UT 84322 USA
25. **Forestry Commission**, Edinburgh, UK
26. Department of Environmental Science and Policy, **George Mason University**, Fairfax, VA 22030 USA
27. Department of Life Sciences, **Natural History Museum**, London SW75BD UK
28. Department of Life Sciences, Silwood Park Campus, **Imperial College London**, Ascot SL57PY UK

ǂ Corresponding author: dougwyu@gmail.com

Supplementary Note 1. Earth Observation Technology

Earth Observation (EO) sensors can be differentiated into active and passive types. Active sensors direct their own source of electromagnetic radiation at the Earth and receive the signal reflected back from the target (e.g. Synthetic Aperture Radar, SAR, transmits microwave pulses). Passive sensors rely on external radiation sources such as the Sun (optical and thermal sensors fall into this category). Different sensors record electromagnetic radiation in specific ranges of the electromagnetic spectrum, with wavelengths from 400-700 nm (visible light) to 700-2400 nm (near to shortwave infrared), 3000-14000 nm (thermally emitted radiation), and 1 cm-1 m (microwave radar wavelengths). Passive EO instruments record radiances at sensor, which generally have to be corrected for atmospheric aerosol and water vapour impacts in order to estimate the land surface reflectances from which EO-derived metrics are usually extracted. Active radar sensors record the transmitted energy that is scattered back from the surface, and since microwaves penetrate clouds, they provide an all-weather observing capability. However, longer wavelengths such as L-band (15-30 cm) and P-band (30-100 cm) can be affected by fluctuations in the total electron content of the ionosphere and the Faraday rotation. Optical and radar sensors are available from both airborne platforms (drones, aircraft) and spaceborne platforms (polar orbiting and geostationary satellites, international space station). Important characteristics of an EO sensor are its spectral coverage and spectral resolution (which bands of the electromagnetic spectrum it measures and at what wavelength detail), its spatial resolution (pixel size), and temporal repeat-frequency (number of days between two acquisitions at the same location). Many applications do not require frequent acquisitions, but multiple images can for instance help account for artefacts and error due to cloud cover[1].

Light Detection and Ranging (LiDAR) is an active remote-sensing technique that transmits infrared or visible polarised light and records the intensity and temporal delay of the received signal. Because of the constant speed of light in air, airborne LiDAR can measure the vertical height of objects with very high accuracy[2]. Radar interferometry from tandem satellite constellations can also measure vertical height but is not as accurate as LiDAR and has a coarser spatial resolution than airborne LiDAR[3]. LiDAR systems can be imaging LiDARs or profiling LiDARs, and some systems record the full waveform of the received radiation, allowing the study of vegetation canopies in great detail, while others only record the first and last return of the waveform. LiDAR instruments are usually mounted on airborne platforms (aircraft, drones) or used as terrestrial instruments (mounted on a tripod or used as a handheld device), with the exception of the spaceborne ICESAT-GLAS profiling LiDAR and the planned GEDI mission to be mounted on the International Space Station.

Supplementary Note 2. Biodiversity and ecosystem information in EO data

The spatial and temporal coverage of EO cannot be matched by *in-situ* surveys, and mapping of habitat extent and land cover types has therefore been incorporated into national EO-monitoring programs for many years[4,5].

Aboveground biomass and carbon storage – Forest ecosystems play a crucial role in global biogeochemical cycles, and deforestation has been a major contributing factor to increasing anthropogenic carbon emissions. Global initiatives such as REDD+ (Reducing emissions from deforestation and forest degradation, and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries) has been negotiated by the UNFCCC for years and was reiterated in the Paris Agreement[6]. While the main aim is to mitigate climate change by reducing carbon emissions, for which developing countries receive results-based payments, safeguards and non-carbon benefits (NCBs) are recognised, including consistency with the conservation of natural forests and biodiversity[7,8]. The success of REDD+ therefore depends on our ability to accurately quantify the global distribution of carbon sources and sinks, for which EO such as SAR or LiDAR are now being developed[9].

Airborne LiDAR can quantify forest canopy height and complexity, and understorey density over large areas, and has been particularly useful in forestry[10]. Although individual trees can be mapped by very high pulse densities[11], forest structure is more commonly described by the heights of a lower density point-cloud aggregated over a forest plot. The average parameters for that forest can then be used to estimate aboveground biomass, which can be translated to ecosystem services like carbon sequestration and storage[12]. Hollaus et al.[13] demonstrated that even simple models could make accurate predictions of timber stock in alpine forests after being calibrated with inventory plot data ($r^2 > 0.80$). The study also showed model accuracy was not sensitive to LiDAR point density or the season of acquisition.

Although performance is likely to vary among habitat types, with accuracy usually greater in low diversity systems, and dependent on the number and size of calibration plots, a meta-analysis of more than 70 studies by Zolkos et al.[14] found airborne LiDAR to be more accurate than radar or passive optical data. Yet more accurate estimates of carbon stocks may be possible using hyperspectral to discriminate tree species[15]. LiDAR can also be used in ecosystems other than forests. For example, Zlinsky et al.[16] demonstrated that LiDAR can replicate ground-based multi-parameter assessments of habitat conservation status in a Natura 2000 grassland reserve in Hungary (Overall Accuracy=0.8); and using EO, the entire reserve could be surveyed.

Biodiversity – While the main focus of REDD+ is to reduce carbon emissions, there is also great potential to improve predictions of spatial patterns of biodiversity from vegetation structure. As argued elsewhere in this paper, these relationships could prove critical to achieving the ambitions of initiatives like REDD+ without compromising the benefits for biodiversity conservation[17].

For instance, early EO products like NDVI (normalised difference vegetation index) have been shown to approximate changes in vegetation structure and hence turnover of the invertebrate ground fauna[18,19], and more recently high spatial resolution airborne imagery has been shown to identify canopy gaps that are associated with the diversity of understorey vegetation[20]. Spectral traits of plants are determined by their physiological and morphological traits, and there are demonstrated applications using EO to reveal the distribution of vegetation types[21,22], functional types[23], richness[24], and temporal changes[25] to name but a few[26]. Nonetheless, the success of habitat mapping varies with habitat type, and research into the right combination of sensors and algorithms is ongoing[27-29]. Finally, the combination of hyperspectral sensors and LiDAR provides an extremely detailed picture of the Earth's surface, potentially capable of identifying the composition of individual trees in some landscapes[30] and reproducing patterns of tree richness and turnover in highly diverse rainforests at landscape scales[22,31,32]. Eventually, similar measurements that directly observe or predict the distribution of biodiversity could be extended globally as satellite-based LiDAR and hyperspectral imaging systems become operational (**S1**).

LiDAR-derived structural metrics have also proven useful as predictors in many *animal* groups[33,34], and LiDAR could be more cost-effective than traditional methods for censusing invertebrate communities[35] and is likely to perform even better once taxonomic uncertainties are reduced with DNA-based identification[36].

Supplementary Note 3. Biodiversity technology

Automated Recording Devices (ARDs)

The first set of technologies encompass ARDs, such as camera traps and bioacoustic recorders that can be left in even remote field locations for weeks to months, capturing records of birds, amphibians, and mammals, and thus allowing continuous sampling of tens of thousands of hectares at a time, with occasional fieldwork to maintain sensors and retrieve data.

Camera traps. - Camera traps are powerful tools for detecting medium to larger-sized mammal species, particularly in forests[37], and they have also been used to study ground-dwelling bird species[38,39] and lizards[40]. Camera traps readily detect rare and cryptic or nocturnal species, and once set up, operate independently of an observer until battery life or memory capacity is exhausted. Early models used film roll cameras and active sensors, where an infrared beam was established across a potential animal path, and the unit was triggered when that beam was broken. Set-up of the infrared beam (height, positioning) had to be tailored specifically to the target species, and early studies often focused on the demography of single charismatic species such as tigers[41,42]. Even with passive heat-in-motion sensors, which made for a more flexible set-up because of the increased area over which animals can be detected, the low number of exposures on film rolls was a severely limiting factor to the time that camera traps could be left in the field without revisiting.

The development of a wide range of digital models in the last 10 years has greatly expanded the applications of camera traps. With increasingly capacious memory cards and batteries, cameras can now routinely be left unattended for weeks up to several months (depending on the expected amount of animal traffic). Options for infrared flash make the equipment nearly invisible, even at night, reducing theft. Modern camera traps capture images of sufficient quality to allow identification to species in 80-90% of photos. Rapid sequential triggers of video options further increase the likelihood of obtaining the footage needed to identify species and individuals. Whereas the up-front investment in the equipment can be high (depending on manufacturer and specifications, a single trap can cost anywhere between $80 and $800), camera traps have repeatedly been shown to beat other methods (e.g. transects, track plates) in their efficiency to document medium to large terrestrial mammal species[43-45], and they become more cost effective for longer surveys[44].

Although the method is still used to study the demography of individual species, particularly those with natural coat patterns allowing individual identification[46-48], camera traps are now also used in behavioural studies[49] and to study species interactions in space and time[50,51]. Moreover, camera traps are increasingly used to survey terrestrial[52-54] and even arboreal mammal communities[55-57]. Camera traps have been proposed as a tool in systematic biodiversity assessments in the context of biodiversity co-benefits of forest management certification and REDD+ payments[58]. As an example application, a recent study on mammalian communities in Bornean forest reserves revealed that particularly threatened species benefit from sustainable forest management practices, applied in the context of certification by the Forest Stewardship Council (FSC)[54]. Similarly, such standardised camera-trapping surveys, if repeated over time, can be used to monitor population and biodiversity trends, which would be impossible using traditional, observer-based fieldwork techniques.

How readily camera traps detect certain species is a function of many factors, including the species' behaviour and abundance, and the specific location and setup of the camera traps[59,60]. For example, arboreal species are harder to detect with ground-based cameras than terrestrial species, and if cameras are set up preferably along roads and trails, species that use these trails will be detected sooner and more frequently than species that prefer to move through vegetation. Comparing biodiversity inventory data across sites and/or years therefore requires a standardised study design, and application of analytical methods that account for these differences in detectability (see Occupancy Modelling, below).

Bioacoustic sensors. - Species that produce acoustic signals can further be surveyed with standalone bioacoustic sensors[61]. Taxonomic groups most frequently studied with bioacoustic methods include birds[62,63], bats[64,65], anurans[63], certain insects[66,67], and cetaceans[68]. Bioacoustic recordings have also been used to study fish[69], and non-flying mammals such as forest elephants[70] and primates[71,72]. Using calls to detect and identify species has a long standing history in bird studies[73]. Handheld sound recorders are a useful tool in such surveys to create permanent records of species audio-detections and to allow for later identification (or verification) of records by specialists. There also exist standalone bioacoustic sensors[74] that, similar to camera traps, can be set up in the field to collect audio information without an observer's presence. Also similar to camera traps, they are primarily limited by battery and storage capacity, and storage capacity has increased dramatically with the switch from analogue to digital equipment[63]. Automatic digital recording systems can be programed to record continuously or at certain times, or, alternatively, more advanced equipment can be triggered by calls above a certain amplitude or of a certain frequency spectrum[75,76].

Once recorded, calls/songs can be identified directly by a trained human observer (but of course only if the species produces a sound that is audible to humans) and/or by visualisation. The latter depicts species-specific acoustic parameters such as the temporal structure and frequency composition of a call/song. Most frequently, visualisation takes the form of a spectrogram, which shows the evolution of the frequency structure of a call over time, using color-coding for changes in amplitude[75]. Such visualisation can reveal call characteristics that the human ear might not perceive. Call-matching to species based on these characteristics can be performed manually, or using computer algorithms. Obrist et al.[75] report that most automated identification software packages achieve a 90% recognition rate but can rarely be expected to cover all species present in a sample. Conversely, Russo and Voigt[65] have voiced concern over the accuracy of automated species identification of bat calls.

Criticism notwithstanding, advances in the development of audio-recorders and call-matching software make automated devices a promising tool for biodiversity inventory and monitoring[58]. Such surveys, however, require extensive preliminary studies to compile reference call data bases. Similar to genetic reference libraries, there are now multiple available sound libraries (e.g. http://www.ibac.info/links.html#libs, accessed 8 Dec 2016), but especially for species-rich tropical communities, bioacoustic databases are currently limited[77]. Circumventing the need for species identification, some studies have suggested the use of bioacoustic diversity as a measure in and of itself. Rather than identifying individual calls and species, this approach is based on measuring the acoustic entropy (i.e. temporal and frequency heterogeneity) of entire soundscapes, and, on the assumption that there is competition for sound niches in time and frequency, a more complex soundscape is taken as an index for a more diverse community[78]. Such bioacoustics diversity indices have been shown to correlate with taxonomic and functional diversity in birds[79] and are a promising field of study, albeit in need of further development and testing[80].

As with other survey methods, detectability and identifiability of individuals and species can be influenced by their vocalisation and other behaviour, habitat, weather, time of day, or the sensitivity of the recording equipment. For example, wind and concurrent vocalisation by other species were found to have a negative impact on the ability to identify frog calls[81], and different equipment has been shown to result in different numbers of bird species detected[82]. In addition to false negatives (i.e. failing to record a species even though it is present), misidentification of calls can result in false positives[83]. As such, standardised surveys and appropriate analytical methods are required to ensure comparability of results across space and time. Occupancy models, for example (discussed below) were developed to account for false negatives and can be adjusted to account for false positives as well[84,85]. They have been successfully used in combination with automated acoustic monitoring[86].

DNA-based methods

Almost all DNA-based techniques exploit the stylised fact that some DNA regions exhibit higher levels of sequence difference between species and low levels of difference within species, which can be used to tell species apart. For animals, the best known of these so-called ‘DNA barcodes’ is a 658-nucleotide portion of the mitochondrial cytochrome oxidase subunit I gene, or COI, which taxonomists have used to build an online reference database that links sequences to species (boldsystems.org, accessed 11 Oct 2016)[87]. Other mitochondrial markers can also be used for taxonomic assignment, and these are available in online databases such as GenBank (blast.ncbi.nlm.nih.gov, accessed 11 Oct 2016). An organism can thus be assigned a taxonomic identification by extracting its DNA, amplifying it with a primer set for the chosen marker(s), sequencing these, and comparing them to a DNA reference database. Even if a species is not represented in a database, its congeners or confamilials usually are, allowing at least higher-level taxonomic identification.

When going from DNA barcoding of single specimens, as described above, to using DNA in synoptic biodiversity surveys, the major challenge is the need to assign taxonomic names to mixed samples containing DNA from multiple taxa, such as occurs in soil, water, faeces, and bulk insect samples. The rise of high-throughput sequencing platforms now makes this routine, and three major approaches are now being used: metabarcoding, high-throughput individual barcoding, and meta/mitogenomics.

Metabarcoding. – DNA is extracted from bulk or environmental samples containing DNA from a mix of different taxa, and a taxonomically informative marker like COI is PCR amplified using a universal primer set targeting the taxonomic group of interest (**Fig. 2** Main Text). In this way, only DNA markers of interest are sequenced, making this a cost-effective approach. The resulting sequences are then clustered into self-similar sets of sequences, each known as an Operational Taxonomic Unit (OTU), which is a species hypothesis. A representative sequence is taken from each OTU and assigned a taxonomy using an online database. The main output of metabarcoding is the classic ecological table of sample X species (OTU), but now achievable for at least hundreds of species across hundreds of samples, plus, to a lesser extent, their phylogenetic relationships. Metabarcoding data thus carry information on species co-occurrence at an unprecedented scale for joint-species-distribution modelling.

Metabarcoding relieves the taxonomic bottleneck, and it also helps relieve the sampling bottleneck. Firstly, metabarcoding can be applied to taxa such as meiofauna and dipterans that are easy to collect and ecologically informative but are so difficult to identify morphologically that they have been ignored in conventional surveys. Secondly, metabarcoding allows difficult-to-find species, such as fungi, fish, and terrestrial vertebrates, to be detected directly from microscopic bits of tissue that can be filtered out of soil, water, air, and parasites, known as 'environmental DNA' or eDNA[88,89]. For instance, leeches, flies, mosquitoes, dung beetles, and ticks retain trace amounts of DNA from their previous meals on animal hosts or faeces, so mass invertebrate trapping could be used to survey other wildlife[90].

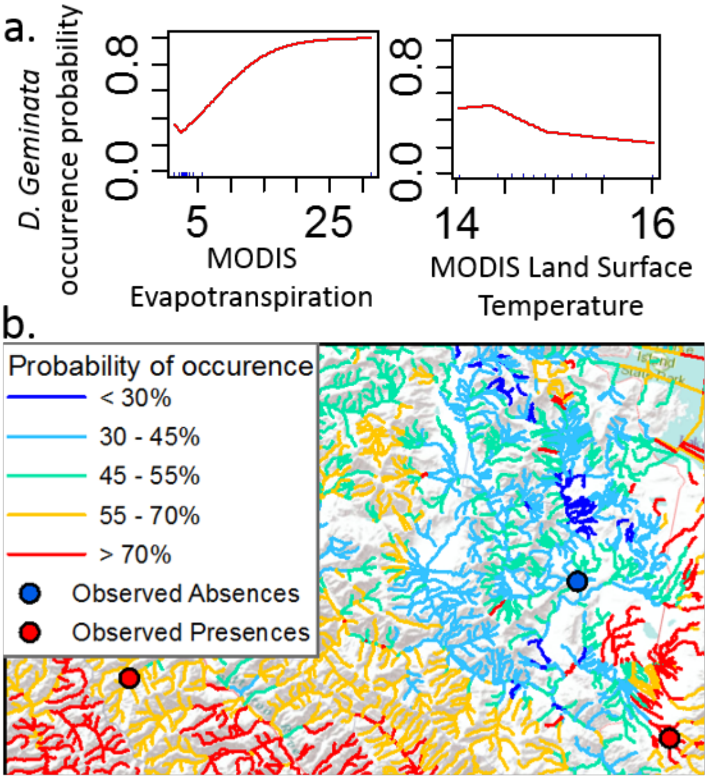
However, metabarcoding unavoidably introduces error, including *inter alia* taxonomic uncertainty due to e.g. PCR and sequencing error and incomplete reference databases, sample cross-contamination, and loss of species, biomass, and abundance information. Judicious sampling and primer design, lab practice, and bioinformatic and statistical pipelines are able to correct or compensate for these errors, and studies have shown that metabarcoding datasets reflect on-the-ground reality sufficiently closely to allow correct management decisions[91-96]. It is worth noting that errors are explicit and quantifiable in DNA-based pipelines, whereas conventional surveys contain important error sources, such as visual misidentifications[97], that are essentially impossible to quantify or correct retrospectively.

High-throughput individual barcoding – In this method[98], large numbers of organisms, typically insects, are *individually* extracted, amplified, and tagged during amplification. Hundreds of individual amplicons are then pooled and sequenced, producing a separate barcode for each organism. Throughput is lower and workload is higher than in metabarcoding, but abundance information is preserved, and individual organisms can be revisited for further taxonomic study.

Meta/mitogenomics – Like metabarcoding, metagenomics can be used on bulk or environmental samples, but instead of targeting a specific gene, all DNA is sequenced, and the output datasets are interrogated *in silico* for taxonomically and functionally informative gene sequences (**Fig. 2** Main Text). Compared to metabarcoding, the advantage of this genomic approach is that it does not require a PCR amplification step to enrich target taxa, which should reduce bias. If samples are sequenced deeply enough, even low-biomass species can be detected in the mix (although sequencer library construction still imposes some biases). Metagenomics also preserves more information on species relative biomasses (a proxy for ecosystem-function importance), can reduce the risk of sample contamination, and depending on the number of samples, can reduce workload. Lastly, it increases the certainty of taxonomic assignment for species that are present in reference database. Currently, metagenomics is routinely applied to microbial communities but is not yet applied to Eukaryotes, due to their much larger genomes and thus higher costs. However, bioinformatic approaches that allow rapid pairwise comparisons of genomic datasets[99] and continued decreases in sequencing costs will make this approach feasible for Eukaryotes. That said, because orders-of-magnitude fewer species have been genome-sequenced, relative to barcode databases, metagenomics applied to Eukaryotes is best suited for studies that focus on hundreds of target species or fewer, for which it is possible to build custom reference databases.

In mitogenomics, the focus is on mitochondrial genomes, which can be individually assembled out of low-coverage sequencing of bulk samples (‘genome skims’), even though mitochondrial reads typically make up <1% of a sequencer’s output[100]. This greatly reduces the cost of building reference databases. Mitogenomics has been used to reconstruct the phylogenetic community structure of soil-dwelling beetle communities[101] and to reliably assign bee species to samples, even after the samples had been DNA-cross-contaminated by handling[102].

Single-species detection – Finally, in situations where it is imperative to detect particular species of concern (e.g. early detection of invasive species or monitoring threatened species) with high probability, older molecular techniques can be used and/or added to the above methods. Species-specific primers can be used in addition to generic primers during metabarcoding to increase detection probability[103], or species-specific quantitative PCR (qPCR) can be conducted on eDNA samples. Although low throughput, this application of targeted qPCR reduces false negatives, provided that proper lab procedure, including negatives controls, is followed[104]. Improved detection rates lead to improvements in model performance, thus increasing the reliability of the predicted distributions of these species of concern, and greater cost efficiency[105]. Single-species detections using qPCR have been combined with MODIS satellite observations to build maximum-entropy species distribution models that predicted the distribution of an invasive diatom (*Didymosphenia geminata*) across the Rocky Mountains (Fig S3.1[106]). Models based on occurrence data from both eDNA and traditional methods correctly predicted occurrence of *D. geminata* at external validation sites with a 93 – 100% correct classification rate (area under the receiver operating characteristic curve, a combined measure of sensitivity and specificity, ranged from 0.94 to 1.00). Temporally concurrent environmental predictors, including evapotranspiration or land surface temperature data from MODIS, allow these models to account for spatial and temporal variation and produce robust predictions (Fig S3.1a). This provides natural resource managers spatially explicit and extensive predictions on where this invasive species is likely to occur. The same approach is also being applied to mapping distributions of six native fish on the north-slope of Alaska to aid in their conservation[106].



Supplementary Figure 3.1 a: Relationship between MODIS measurements of evapotranspiration and land surface temperature and the occurrence of the invasive diatom D. geminata. b: Resulting maps of probability of D. geminata occurrence from applying model to individual stream segments.

Supplementary Note 4. Statistical modelling

Occupancy Detection Models – Logistical constraints dictate that a site-by-species matrix can only ever comprise a finite set of point samples, leaving most of the environment unsampled. Moreover, even within sampled sites, an unavoidable problem is false negatives: species that are indeed present but not detected[107,108], and in some cases false positives (species detected that are in fact absent). To correct for imperfect detection, occupancy-detection models are used to disentangle the factors that determine the occurrence of a species from those that affect the probability of detection, given occurrence[109]. To estimate the probability of detection, a location is repeatedly sampled, either by spatially sub-sampling a site, or by re-visiting the same location multiple times within a short time period. A hierarchical generalised linear mixed model (GLMM) – technically a zero-inflated logistic regression of species detection/non-detection data – is then used to predict the probability that a species occurs at a site, based on the site's environmental covariates and the empirically estimated probability of detection, which can also itself be a function of site- and time-specific covariates.

Community Occupancy Detection Models – In the simplest application of occupancy detection, each species is considered independent, so a multi-species model simply combines the species' environmental responses and their different detectabilities, and calculates metrics of diversity either from occupancy probabilities (in a likelihood framework, richness is the sum of all occupancy probabilities at a site), or from realised occupancy states (in a Bayesian framework, richness is the number of species estimated to occur at that site)[110,111]. However, if the environmental responses of multiple species follow a common distribution, community occupancy detection models allow individual coefficients to be modelled as a random effect, whereby the data-poor species borrow information from data-rich species[112,113], which allows information on species traits to be included as predictors[114]. Furthermore, based on differences in species detection probabilities, occupancy models can also estimate the number of species that were never detected, by introducing zero-inflation within the inputs ("data augmentation"[115,116]), recently extended for multi-region comparison[117]. More complex models can include the effect of community dynamics on spatial and temporal variation in occurrence[118].

Joint Species Distribution Models / Latent Variable Models – An extension of the single-species approach is to consider all pairwise co-occurrences among species[119]. These so-called joint species distribution models (JSDMs) predict multi-species responses by not only modelling species-specific responses to environmental covariates as random effects but also accounting for residual patterns of co-occurrence not explained by environmental factors[120]. In the past, the number of taxa that J-SDMs could consider was limited because the number of parameters in unstructured variance-covariance matrices rises rapidly[121]. However, JSDMs can now analyse high numbers of species by inducing correlation among taxa using ‘latent’ unobserved factors[120]. Residual correlation might indicate species interactions, like competition or predation, unmeasured predictors, spatial autocorrelation, or misspecification of the model, all of which warrant further investigation[122]. Spatially explicit latent variables allow one to predict a species community for a focal site using as predictors not only the environmental variables measured at the focal site, but also the occurrences and co-occurrences of the species in nearby sites[123], thus providing a statistically efficient tool for producing interpolated species distribution maps from sparse data on species rich communities. In principle, detection probability itself could also be included as a layer describing the observation process[124-126]. LVMs are currently an area of active research, and there has been rapid progress to expand computational limits and integrate with the breadth of previous development using hierarchical mixed models[120]. Of particular interest is the opportunity to cluster species responses to environmental covariates according to species traits (i.e. “the fourth-corner problem”[127]) making it easier to translate compositional turnover to functional shifts[128].

Generalised Dissimilarity Models – Finally, in very diverse communities with hundreds or thousands of taxa (e.g. soil fauna), it might not be meaningful to model the responses of individual species. Instead, generalised dissimilarity models (GDM) use a pairwise matrix of compositional dissimilarity to predict the nonlinear response of compositional turnover to environmental changes; weighting and transforming environmental variables so that conversion of multidimensional environmental space best describes the scaled turnover of biological composition[129]. GDM can help identify new sampling sites for more reliable prediction[130], and uncertainty in variable selection can be further evaluated using Bayesian bootstrapping[131]. The dissimilarity matrix can also be derived from other biological distance metrics like sequence reads, allelic turnover, functional differences, or phylogenetic diversity[99,132,133]. The link between turnover of composition or function can then be tested using scaled environmental variables as predictors of spatial or temporal changes in service provision[134].

GDM has already been incorporated into EO-based applications to estimate ecological values at landscape scales[135,136], and model performance improves when combined with multispectral EO sensors[137]. By predicting the dissimilarity of sites alongside an expected species-area relationship, GDM can also be used to estimate the proportion of biodiversity retained regionally (**Box 3** in main text). This has numerous conservation applications (e.g. protected areas effectiveness – Aichi Target 11[138]), as well as quantifying the biodiversity left regionally (gamma diversity) to support ecosystem services[139]. If the identity of species composition is still desired, GDM can be combined with a model of alpha diversity to estimate the probable species composition of every cell in a landscape[140]. Furthermore, ecological processes like dispersal, growth rates, and metacommunity dynamics have been incorporated to predict ecosystem function and to rank management actions[134,141].

Sampling design. – For a given sampling effort, careful survey design can improve the accuracy and reliability of biodiversity models. For example, although we may be interested in the different species assemblages that each contribute to carbon storage in forests (e.g. large frugivorous mammals, isopods, and saprotrophic fungi; **Box 2** in main text**)**, we would not sample for these disparate taxa at the same spatial grain[142]. For example, within the home range of a single monkey troupe, the composition of saprotrophic fungi might exhibit high levels of turnover across wood from different tree species. Fungi thus need to be sampled at a finer spatial grain than do mammals, and soil fauna might need to be sampled even more finely again. The grain of sampling should therefore try to match the grain at which environmental heterogeneity creates different habitats, but for efficiency, should also aggregate across points whose composition only varies due to stochastic fluctuations. Once the scale of analysis has been chosen, the sampling strategy can be adjusted to ensure that biological samples represent an equivalent area effectively. For example, how many soil cores need to be sequenced to estimate the diversity of fungi within a forest plot? Scaling relationships like species-area curves can be a useful shortcut to compare community-level characteristics (alpha or beta diversity), but they cannot identify which species were missed. Instead, the differences in sampling effort, including the area sampled (e.g. number of quadrats), could be accounted for explicitly by species' detectability in community occupancy models[109]. Finally, it is important that samples capture the full range and variability of environments across the region of study, especially where environmental differences lead to higher turnover[130]. This is particularly true of finite resources that become limiting, such as soil moisture gradients that determine vegetation succession in arid biomes, but are less important to predicting turnover in the wet tropics. These decisions can be guided by expert opinion and existing survey data, but pilot studies may be required at multiple spatial and temporal resolutions, before settling on a single strategy.

In addition to the pathways that we have described above and in the Main Text (Figure 3), there of course exist other methods to model communities, which take into explicit account biological mechanisms such as demography, dispersal, evolution, and specialist interactions[143,144]. We have not covered these methods because they require much more input data[145] and thus are limited in their applicability, although when available, all information should of course be exploited. We reiterate that species co-occurrence matrices, latent variables, phylogenetic structure, and ecological functions can all be extracted from the three statistical pathways in Figure 3, and these provide an efficient way to generate causal hypotheses for further, targeted investigation.

Acknowledgments for individual authors

D. W. Yu was supported by the National Natural Science Foundation of China (31400470, 41661144002, 31670536, 31500305, GYHZ1754), the Ministry of Science and Technology of China (2012FY110800), the University of East Anglia, and the State Key

Laboratory of Genetic Resources and Evolution at the Kunming Institute of Zoology (GREKF13-13, GREKF14-13, GREKF16-09).

M. T. P. Gilbert was supported by Danish National Research Foundation grant DNRF94 and Lundbeck Foundation grant R52-5062.

A. Wilting was supported by the German Federal Ministry of Education and Research grant BMBF FKZ: 01LN1301A.

L. Matthews, C. A. Cobbold, and R. Reeve were supported by BBSRC grant BB/N013336/1.

A. P. Vogler was supported by the Biodiversity Initiative of the Natural History Museum. The Initiative funded an international symposium ‘Tropical Biodiversity in the 21st Century’ (June 2013) that convened researchers in tropical biology, EO and molecular taxonomy.

A. Zlinszky was funded by the PD 115833 grant of the Hungarian Science Fund (OTKA)

M. Herold was supported by the Global Forest Observation Initiative (GFOI) research and development programme and the Global Observations of Forest Cover and Land Dynamics (GOFC-GOLD) project office funded by the European Space Agency.

C. Martius and M. Herold were supported by the Norwegian Agency for Development Cooperation (Norad) and the International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) with financial support from the donors to the CGIAR Fund. This work of M. Herold was further supported by the Global Forest Observation Initiative (GFOI) research and development programme and the sponsor of Global Observations of Forest Cover and Land Dynamics (GOFC-GOLD). Kim Calders and Jose Gonzales are acknowledged for providing the TLS data for the Elephant in Gabon.

B. C. Emerson was supported by Spanish MINECO grant CGL2013-42589-P, co-financed by FEDER, and Spanish MAGRAMA grant S20141203_002597 from the Organismo Autonomo Parques Nacionales.

J. R. Olson and T. W. Rodgers are supported by NASA grant NNX14AC40G and NNX15AR79G.

K. Bohmann was supported by The Danish Council for Independent Research grant DFF – 5051-00140.

M.J. Wooster was supported by the NERC National Center for Earth Observation (NCEO), UK.

B. Cole was supported by the NERC Biodiversity and Ecosystem Services Sustainability (BESS) program.

H. Balzter was supported by the NERC Biodiversity and Ecosystem Services Sustainability (BESS) program and the NERC National Center for Earth Observation (NCEO), UK.

References

1 Wilson, A. M. & Jetz, W. Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. *PLOS Biology* **14**, e1002415, doi:10.1371/journal.pbio.1002415 (2016).

2 Bradbury, R. B. *et al.* Modelling relationships between birds and vegetation structure using airborne LiDAR data: a review with case studies from agricultural and woodland environments. *Ibis* **147**, 443-452, doi:10.1111/j.1474-919x.2005.00438.x (2005).

3 Balzter, H., Baade, J. & Rogers, K. Validation of the TanDEM-X Intermediate Digital Elevation Model With Airborne LiDAR and Differential GNSS in Kruger National Park.

IEEE Geoscience and Remote Sensing Letters **13**, 277-281, doi:10.1109/LGRS.2015.2509500 (2016).

4 Firbank, L. G. *et al.* Assessing stock and change in land cover and biodiversity in GB: an introduction to Countryside Survey 2000. *Journal of Environmental Management* **67**, 207-218, doi:10.1016/S0301-4797(02)00174-3 (2003).

5 Duro, D. C., Coops, N. C., Wulder, M. A. & Han, T. Development of a large area biodiversity monitoring system driven by remote sensing. *Progress in Physical Geography* **31**, 235-260, doi:10.1177/0309133307079054 (2007).

6 UNFCCC. Adoption of the Paris Agreement. FCCC/CP/2015/L.9/Rev.1. (2015).

7 UNFCCC. Report of the Conference of the Parties on its sixteenth session, held in Cancun from 29 November to 1- December 2010. FCCC/CP/2010/7/Add.1. Decision 1/CP.16., (2010).

8 Turnhout, E. *et al.* Envisioning REDD+ in a post-Paris era: between evolving expectations and current practice. *WIREs Clim Change* **8**, e425, doi:doi:10.1002/wcc.425 (2016).

9 Lynch, J., Maslin, M., Balzter, H. & Sweeting, M. Sustainability: Choose satellites to monitor deforestation. *Nature* **496**, 293-294, doi:10.1038/496293a (2013).

10 Vierling, K. T., Vierling, L. A., Gould, W. A., Martinuzzi, S. & Clawges, R. M. Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment* **6**, 90-98, doi:10.1890/070001 (2008).

11 Maltamo, M., Mustonen, K., Hyyppä, J., Pitkänen, J. & Yu, X. The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. *Canadian Journal of Forest Research* **34**, 1791-1801, doi:10.1139/x04-055 (2004).

12 McKinley, D. C. *et al.* A synthesis of current knowledge on forests and carbon storage in the United States. *Ecological Applications* **21**, 1902-1924, doi:10.1890/10-0697.1 (2011).

13 Hollaus, M., Wagner, W., Schadauer, K., Maier, B. & Gabler, K. Growing stock estimation for alpine forests in Austria: a robust lidar-based approach. *Canadian Journal of Forest Research* **39**, 1387-1400, doi:10.1139/X09-042 (2009).

14 Zolkos, S. G., Goetz, S. J. & Dubayah, R. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment* **128**, 289-298, doi:10.1016/j.rse.2012.10.017 (2013).

15 Dalponte, M. & Coomes, D. A. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution* **7**, 1236–1245, doi:10.1111/2041-210X.12575 (2016).

16 Zlinszky, A., Deák, B., Kania, A., Schroiff, A. & Pfeifer, N. Mapping Natura 2000 Habitat Conservation Status in a Pannonic Salt Steppe with Airborne Laser Scanning. *Remote Sensing* **7**, 2991-3019, doi:10.3390/rs70302991 (2015).

17 Beaudrot, L. *et al.* Limited carbon and biodiversity co-benefits for tropical forest mammals and birds. *Ecological Applications* **26**, 1098-1111, doi:10.1890/15-0935 (2016).

18 Lassau, S. A., Cassis, G., Flemons, P. K. J., Wilkie, L. & Hochuli, D. F. Using high-resolution multi-spectral imagery to estimate habitat complexity in open-canopy forests: can we predict ant community patterns? *Ecography* **28**, 495-504, doi:10.1111/j.0906-7590.2005.04116.x (2005).

19 Lassau, S. A. & Hochuli, D. F. Testing predictions of beetle community patterns derived empirically using remote sensing. *Diversity and Distributions* **14**, 138-147, doi:10.1111/j.1472-4642.2007.00438.x (2008).

20 Getzin, S., Wiegand, K. & Schöning, I. Assessing biodiversity in forests using very high-resolution images and unmanned aerial vehicles. *Methods in Ecology and Evolution* **3**, 397-404, doi:10.1111/j.2041-210X.2011.00158.x (2012).

21 Gillespie, T. W., Foody, G. M., Rocchini, D., Giorgi, A. P. & Saatchi, S. Measuring and modelling biodiversity from space. *Progress in Physical Geography* **32**, 203-221, doi:10.1177/0309133308093606 (2008).

22 Asner, G. P. *et al.* Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science* **355**, 385-389, doi:10.1126/science.aaj1987 (2017).

23 Ustin, S. L. & Gamon, J. A. Remote sensing of plant functional types. *New Phytologist* **186**, 795-816, doi:10.1111/j.1469-8137.2010.03284.x (2010).

24 Fricker, G. A., Wolf, J. A., Saatchi, S. S. & Gillespie, T. W. Predicting spatial variations of tree species richness in tropical forests from high-resolution remote sensing. *Ecological Applications* **25**, 1776-1789, doi:10.1890/14-1593.1 (2015).

25 Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850-853, doi:10.1126/science.1244693 (2013).

26 Lausch, A. *et al.* Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. *Ecological Indicators* **70**, 317-339, doi:10.1016/j.ecolind.2016.06.022 (2016).

27 Pfeifer, M., Disney, M., Quaife, T. & Marchant, R. Terrestrial ecosystems from space: a review of earth observation products for macroecology applications. *Global Ecology and Biogeography* **21**, 603-624, doi:10.1111/j.1466-8238.2011.00712.x (2012).

28 Petrou, Z. I., Manakos, I. & Stathaki, T. Remote sensing for biodiversity monitoring: a review of methods for biodiversity indicator extraction and assessment of progress

towards international targets. *Biodiversity and Conservation* **24**, 2333-2363, doi:10.1007/s10531-015-0947-z (2015).

29 Rocchini, D. *et al.* Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sensing in Ecology and Conservation* **2**, 25-36, doi:10.1002/rse2.9 (2016).

30 Colgan, M. S., Baldeck, C. A., Féret, J.-B. & Asner, G. P. Mapping Savanna Tree Species at Ecosystem Scales Using Support Vector Machine Classification and BRDF Correction on Airborne Hyperspectral and LiDAR Data. *Remote Sensing* **4**, 3462-3480, doi:10.3390/rs4113462 (2012).

31 Féret, J.-B. & Asner, G. P. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Applications* **24**, 1289-1296, doi:10.1890/13-1824.1 (2014).

32 Féret, J.-B. & Asner, G. P. Microtopographic controls on lowland Amazonian canopy diversity from imaging spectroscopy. *Ecological Applications* **24**, 1297-1310, doi:10.1890/13-1896.1 (2014).

33 Davies, A. B. & Asner, G. P. Advances in animal ecology from 3D-LiDAR ecosystem mapping. *Trends in Ecology & Evolution* **29**, 681691, doi:10.1016/j.tree.2014.10.005 (2014).

34 Simonson, W. D., Allen, H. D. & Coomes, D. A. Applications of airborne lidar for the assessment of animal species diversity. *Methods in Ecology and Evolution* **5**, 719729, doi:10.1111/2041-210X.12219 (2014).

35 Müller, J. & Brandl, R. Assessing biodiversity by remote sensing in mountainous terrain: the potential of LiDAR to predict forest beetle assemblages. *Journal of Applied Ecology* **46**, 897-905, doi:10.1111/j.1365-2664.2009.01677.x (2009).

36 Vierling, K. T. *et al.* Spinning a laser web: predicting spider distributions using LiDAR. *Ecological Applications* **21**, 577–588, doi:10.1890/09-2155.1 (2011).

37 Sunarto *et al.* Threatened predator on the equator: multi-point abundance estimates of the tiger Panthera tigris in central Sumatra. *Oryx* **47**, 211-220, doi:10.1017/S0030605311001530 (2013).

38 O'Brien, T. G. & Kinnaird, M. F. A picture is worth a thousand words: the application of camera trapping to the study of birds. *Bird Conserv. Int. [Internet]* **18** (2008).

39 Suwanrat, S., Ngoprasert, D., Sutherland, C., Suwanwaree, P. & Savini, T. Suwanrat S, Ngoprasert D, Sutherland C, Suwanwaree P, Savini T. 2015. Estimating density of secretive terrestrial birds (Siamese Fireback) in pristine and degraded forest using camera traps and distance sampling. *Global Ecology and Conservation* **3**, 596-606 (2015).

40 Broeckhoven, C. & le Fras Nortier Mouton, P. Some Like It Hot: Camera Traps Unravel the Effects of Weather Conditions and Predator Presence on the Activity Levels of Two Lizards. *PLoS One* **10:e0137428** (2015).

41 Karanth, K. U. & Nichols, J. D. Estimation of tiger densities in India using photographic captures and recaptures. *Ecology* **79**, 2852–2862 (1998).

42 Karanth, K. U., Nichols, J. D., Kumar, N. S. & Hines, J. E. Assessing tiger population dynamics using photographic capture-recapture sampling. *Ecology* **87**, 2925–2937 (2006).

43 Silveira, L., Jácomo, A. T. A. & Diniz-Filho, J. A. F. Camera trap, line transect census and track surveys: a comparative evaluation. *Biological Conservation*, 351–355 (2003).

44 Lyra-Jorge, M. C., Ciocheti, G., Pivello, V. R. & Meirelles, S. T. Comparing methods for sampling large- and medium-sized mammals: camera traps and track plots. *European Journal of Wildlife Research* **54**, 739–744 (2008).

45 Balme, G. A., Hunter, L. T. B. & Slotow, R. Evaluating Methods for Counting Cryptic Carnivores. *Journal of Wildlife Management* **73**, 433–441 (2009).

46 Gardner, B., Reppucci, J., Lucherini, M. & Royle, J. A. Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. *Ecology* **91**, 3376–3383 (2010).

47 Sollmann, R. *et al.* Improving density estimates for elusive carnivores: Accounting for sex-specific detection and movements using spatial capture–recapture models for jaguars in central Brazil. *Biological Conservation* **144**, 1017-1024, doi:10.1016/j.biocon.2010.12.011 (2011).

48 Wilting, A. *et al.* Density of the Vulnerable Sunda clouded leopard Neofelis diardi in two commercial forest reserves in Sabah, Malaysian Borneo. *Oryx* **46**, 423-426, doi:10.1017/S0030605311001694 (2012).

49 Armenteros, J. A. *et al.* Do wild Red-legged Partridges (*Alectoris rufa*) use feeders? An investigation of their feeding patterns using camera trapping. *Avian Biology Research* **8**, 14–24 (2015).

50 Linkie, M. & Ridout, M. S. Assessing tiger–prey interactions in Sumatran rainforests. *Journal of Zoology* **284**, 224-229, doi:10.1111/j.1469-7998.2011.00801.x (2011).

51 Sollmann, R. *et al.* Using occupancy models to investigate space partitioning between two sympatric large predators, the jaguar and puma in central Brazil. *Mammalian Biology - Zeitschrift für Säugetierkunde* **77**, 41-46, doi:10.1016/j.mambio.2011.06.011 (2012).

52 Beaudrot, L. *et al.* Standardized Assessment of Biodiversity Trends in Tropical Forest Protected Areas: The End Is Not in Sight. *PLoS Biol* **14**, e1002357, doi:10.1371/journal.pbio.1002357 (2016).

53 Brodie, J. F. *et al.* Correlation and persistence of hunting and logging impacts on tropical rainforest mammals. *Conservation Biology* **29**, 110-121, doi:10.1111/cobi.12389 (2015).

54 Sollmann, R. *et al.* Quantifying mammal biodiversity co-benefits in certified tropical forests. *Diversity and Distributions* **23**, 317-328, doi:10.1111/ddi.12530 (2017).

55 Gregory, T., Carrasco Rueda, F., Deichmann, J., Kolowski, J. & Alonso, A. Arboreal camera trapping: taking a proven method to new heights. *Methods in Ecology and Evolution* **5**, 443-451, doi:10.1111/2041-210X.12177 (2014).

56 Bowler, M. T., Tobler, M. W., Endress, B. A., Gilmore, M. P. & Anderson, M. J. Estimating mammalian species richness and occupancy in tropical forest canopies with arboreal camera traps. *Remote Sensing in Ecology and Conservation*, doi:10.1002/rse2.35 (2016).

57 Whitworth, A., Braunholtz, L. D., Huarcaya, R. P., MacLeod, R. & Beirne, C. Out on a Limb: Arboreal Camera Traps as an Emerging Methodology for Inventorying Elusive Rainforest Mammals. *Tropical Conservation Science* **9**, 675–698 (2016).

58 Waldon, J., Miller, B. W. & Miller, C. M. A Model Biodiversity Monitoring Protocol for REDD Projects. *Tropical Conservation Science* **4**, 254-260 (2011).

59 Harmsen, B. J., Foster, R. J., Silver, S., Ostro, L. & Doncaster, C. P. Differential Use of Trails by Forest Mammals and the Implications for Camera-Trap Studies: A Case Study from Belize. *Biotropica* **42**, 126-133, doi:10.1111/j.1744-7429.2009.00544.x (2010).

60 Sollmann, R., Mohamed, A., Samejima, H. & Wilting, A. Risky business or simple solution – Relative abundance indices from camera-trapping. *Biological Conservation* **159**, 405–412 (2013).

61 Blumstein, D. T. *et al.* Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology* **48**, 758-767, doi:10.1111/j.1365-2664.2011.01993.x (2011).

62 Hobson, K. A., Rempel, R. S., Greenwood, H., Turnbull, B. & Van Wilgenburg, L. S. Acoustic surveys of birds using electronic recordings: New potential from an omnidirectional microphone system. *Wildlife Society Bulletin* **30**, 709-720, doi:10.2307/3784223 (2002).

63 Acevedo, M. A. & Villanueva-Rivera, L. J. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* **34**, 211-214, doi:10.2193/0091-7648(2006)34[211:UADRSA]2.0.CO;2 (2006).

64 O'Farrell, M. J. & Gannon, W. L. A Comparison of Acoustic versus Capture Techniques for the Inventory of Bats. *Journal of Mammalogy* **80**, 24-30, doi:10.2307/1383204 (1999).

65 Russo, D. & Voigt, C. C. The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for a sound analysis. *Ecological Indicators* **66**, 598-602, doi:10.1016/j.ecolind.2016.02.036 (2016).

66 Diwakar, S. & Balakrishnan, R. The assemblage of acoustically communicating crickets of a tropical evergreen forest i southern India: call diversity and diel calling patterns. *Bioacoustics* **16**, 113-135, doi:10.1080/09524622.2007.9753571 (2007).

67 Lehmann, G. U. C., Frommolt, K.-H., Lehmann, A. W. & Riede, K. Baseline data for automated acoustic monitoring of Orthoptera in a Mediterranean landscape, the

Hymettos, Greece. *Journal of Insect Conservation* **18**, 909-925, doi:10.1007/s10841-014-9700-2 (2014).

68 Sousa-Lima, R. S., Fernandes, D. P., Norris, T. F. & Oswald, J. N. A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals: 2013 state-of-the-industry. *IEEE/OES Acoustics in Underwater Geosciences Symposium, Rio de Janeiro, 2013*, 1-9, doi:doi: 10.1109/RIOAcoustics.2013.6683984 (2013).

69 Rountree, R. A. *et al.* Listening to Fish. *Fisheries* **31**, 433-446, doi:10.1577/1548-8446(2006)31[433:LTF]2.0.CO;2 (2006).

70 Thompson, M. E., Schwager, S. J., Payne, K. B. & Turkalo, A. K. Acoustic estimation of wildlife abundance: methodology for vocal mammals in forested habitats. *African Journal of Ecology* **48**, 654-661, doi:10.1111/j.1365-2028.2009.01161.x (2010).

71 Heinicke, S. *et al.* Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution* **6**, 753-763, doi:10.1111/2041-210X.12384 (2015).

72 Kalan, A. K. *et al.* Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators* **54**, 217-226, doi:10.1016/j.ecolind.2015.02.023 (2015).

73 Graber, R. R. & Cochran, W. W. An Audio Technique for the Study of Nocturnal Migration of Birds. *The Wilson Bulletin* **71**, 220-236 (1959).

74 Peterson, M. E. & Dorcas, C. R. in *Measuring and Monitoring Biological Diversity: Standard Methods for Amphibians* (eds W.R. Heyer *et al.*) 47-57 (Smithsonian Institution, 1994).

75 Obrist, M. K., Pavan, G. J. S., Riede, K., Llusia, D. & Márquez, R. in *Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories. Vol. 8. ABC Taxa.* (eds J. Eymann *et al.*) 68-99 (Belgian Development Cooperation, 2010).

76 Stahlschmidt, P. & Brühl, C. A. Bats as bioindicators – the need of a standardized method for acoustic bat activity surveys. *Methods in Ecology and Evolution* **3**, 503-508, doi:10.1111/j.2041-210X.2012.00188.x (2012).

77 Walters, C. L. *et al.* in *Bat Evolution, Ecology, and Conservation* (eds Rick A. Adams & Scott C. Pedersen) 479-499 (Springer New York, 2013).

78 Sueur, J., Pavoine, S., Hamerlynck, O. & Duvail, S. Rapid Acoustic Survey for Biodiversity Appraisal. *PLOS ONE* **3**, e4065, doi:10.1371/journal.pone.0004065 (2009).

79 Gasc, A. *et al.* Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? *Ecological Indicators* **25**, 279-287, doi:10.1016/j.ecolind.2012.10.009 (2013).

80 Sueur, J., Farina, A., Gasc, A., Pieretti, N. & Pavoine, S. Acoustic Indices for Biodiversity Assessment and Landscape Investigation. *Acta Acustica united with Acustica* **100**, 772-781, doi:10.3813/AAA.918757 (2014).

81 Aide, T. M. *et al.* Real-time bioacoustics monitoring and automated species identification. *PeerJ* **1**, e103, doi:10.7717/peerj.103 (2013).

82 Rempel, R. S., Francis, C. M., Robinson, J. N. & Campbell, M. Comparison of audio recording system performance for detecting and monitoring songbirds. *Journal of Field Ornithology* **84**, 86-97, doi:10.1111/jofo.12008 (2013).

83 Towsey, M., Planitz, B., Nantes, A., Wimmer, J. & Roe, P. A toolbox for animal call recognition. *Bioacoustics* **21**, 107-125, doi:10.1080/09524622.2011.648753 (2012).

84 Miller, D. A. *et al.* Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology* **92**, 1422-1428, doi:10.1890/10-1396.1 (2011).

85 Miller, D. A. W. *et al.* Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecological Applications* **22**, 1665-1674, doi:10.1890/11-2129.1 (2012).

86 Campos-Cerqueira, M. & Aide, T. M. Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. *Methods in Ecology and Evolution* **7**, 1340-1348, doi:10.1111/2041-210X.12599 (2016).

87 Ratnasingham, S. & Hebert, P. D. N. bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes* **7**, 355-364, doi:10.1111/j.1471-8286.2007.01678.x (2007).

88 Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution* **29**, 358-367, doi:10.1016/j.tree.2014.04.003 (2014).

89 Thomsen, P. F. & Willerslev, E. Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**, 4-18, doi:10.1016/j.biocon.2014.11.019 (2015).

90 Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P. & Schubert, G. An invertebrate stomach's view on vertebrate ecology. *BioEssays* **35**, 1004-1013, doi:10.1002/bies.201300060 (2013).

91 Ji, Y. *et al.* Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* **16**, 1245-1257, doi:10.1111/ele.12162 (2013).

92 Lejzerowicz, F. *et al.* High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Scientific Reports* **5**, 13932, doi:10.1038/srep13932 (2015).

93 Edwards, D. P. *et al.* Selective-logging and oil palm: multitaxon impacts, biodiversity indicators, and trade-offs for conservation planning. *Ecological Applications* **24**, 2029-2049, doi:10.1890/14-0010.1 (2014).

94 Gibson, J. F. *et al.* Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLoS ONE* **10**, e0138432, doi:10.1371/journal.pone.0138432 (2015).

95 Hänfling, B. *et al.* Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology* **25**, 3101-3119, doi:10.1111/mec.13660 (2016).

96 Madden, A. A. *et al.* The diversity of arthropods in homes across the United States as determined by environmental DNA analyses. *Molecular Ecology* **25**, 6214-6224, doi:10.1111/mec.13900 (2016).

97 Austen, G. E., Bindemann, M., Griffiths, R. A. & Roberts, D. L. Species identification by experts and non-experts: comparing images from field guides. *Scientific Reports* **6**, 33634, doi:10.1038/srep33634 (2016).

98 Meier, R., Wong, W., Srivathsan, A. & Foo, M. $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. *Cladistics* **32**, 100-110, doi:10.1111/cla.12115 (2016).

99 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* **17**, 132, doi:10.1186/s13059-016-0997-x (2016).

100 Crampton-Platt, A., Yu, D. W., Zhou, X. & Vogler, A. P. Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience* **5**, 1-11, doi:10.1186/s13742-016-0120-y (2016).

101 Andújar, C. *et al.* Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology* **24**, 3603-3617, doi:10.1111/mec.13195 (2015).

102 Tang, M. *et al.* High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution* **6**, 1034-1043, doi:10.1111/2041-210X.12416 (2015).

103 Schubert, G. *et al.* Targeted detection of mammalian species using carrion fly-derived DNA. *Molecular Ecology Resources* **15**, 285-294, doi:10.1111/1755-0998.12306 (2015).

104 Wilcox, T. M. *et al.* Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char *Salvelinus fontinalis*. *Biological Conservation* **194**, 209-216, doi:10.1016/j.biocon.2015.12.023 (2016).

105 Lahoz-Monfort, J. J., Guillera-Arroita, G. & Tingley, R. Statistical approaches to account for false-positive errors in environmental DNA samples. *Molecular Ecology Resources* **16**, 673-685, doi:10.1111/1755-0998.12486 (2016).

106 Olson, J. R., Hawkins, C. P., Mock, K., Huntington, J. & Susfalk., R. System for Mapping And Predicting Species Of Concern (SMAP-SOC), Phase I Final Report and Phase II Plan., (NASA Earth Science Division/Applied Sciences Program, Washington D.C., 2014).

107 Chen, G., Kéry, M., Plattner, M., Ma, K. & Gardner, B. Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology* **101**, 183-191, doi:10.1111/1365-2745.12021 (2013).

108 Tingley, M. W. & Beissinger, S. R. Cryptic loss of montane avian richness and high community turnover over 100 years. *Ecology* **94**, 598-609, doi:10.1890/12-0928.1 (2013).

109 Guillera-Arroita, G. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* **40**, 281-295, doi:10.1111/ecog.02445 (2017).

110 Dorazio, R. M. & Royle, J. A. Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association* **100**, 389-398 (2005).

111 Mihaljevic, J. R., Joseph, M. B. & Johnson, P. T. J. Using multispecies occupancy models to improve the characterization and understanding of metacommunity structure. *Ecology* **96**, 1783-1792, doi:10.1890/14-1580.1 (2015).

112 Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models*. (Cambridge University Press, 2006).

113 Ovaskainen, O. & Soininen, J. Making more out of sparse data: hierarchical modeling of species communities. *Ecology* **92**, 289-295, doi:10.1890/10-1251.1 (2011).

114 Pollock, L. J., Morris, W. K. & Vesk, P. A. The role of functional traits in species distributions revealed through a hierarchical model. *Ecography* **35**, 716-725, doi:10.1111/j.1600-0587.2011.07085.x (2012).

115 Royle, A. J., Dorazio, R. M. & Link, W. A. Analysis of Multinomial Models With Unknown Index Using Data Augmentation. *Journal of Computational and Graphical Statistics* **16**, 67-85, doi:10.1198/106186007X181425 (2007).

116 Royle, A. J. & Dorazio, R. M. Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. *Journal of Ornithology* **152**, 521-537, doi:10.1007/s10336-010-0619-4 (2012).

117 Sutherland, C., Brambilla, M., Pedrini, P. & Tenan, S. A multiregion community model for inference about geographic variation in species richness. *Methods in Ecology and Evolution* **7**, 783-791, doi:10.1111/2041-210X.12536 (2016).

118 Dorazio, R. M., Kéry, M., Royle, J. A. & Plattner, M. Models for inference in dynamic metacommunity systems. *Ecology* **91**, 2466-2475, doi:10.1890/09-1033.1 (2010).

119 Pollock, L. J. *et al.* Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution* **5**, 397-406, doi:10.1111/2041-210X.12180 (2014).

120 Warton, D. I. *et al.* So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution* **30**, 766-779, doi:10.1016/j.tree.2015.09.007 (2015).

121 Ovaskainen, O., Hottola, J. & Siitonen, J. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* **91**, 2514-2521, doi:10.1890/10-0173.1 (2010).

122 Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution* **7**, 549-555, doi:10.1111/2041-210X.12501 (2016).

123 Ovaskainen, O., Roy, D. B., Fox, R. & Anderson, B. J. Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution* **7**, 428-436, doi:10.1111/2041-210X.12502 (2016).

124 Dorazio, R. M., Connor, E. F. & Askins, R. A. Estimating the Effects of Habitat and Biological Interactions in an Avian Community. *PLoS ONE* **10**, e0135987, doi:10.1371/journal.pone.0135987 (2015).

125 Beissinger, S. R. *et al.* Incorporating Imperfect Detection into Joint Models of Communities: A response to Warton et al. *Trends in Ecology & Evolution* **31**, 736-737, doi:10.1016/j.tree.2016.07.009 (2016).

126 Warton, D. I. *et al.* Extending Joint Models in Community Ecology: A Response to Beissinger et al. *Trends in Ecology & Evolution* **31**, 737–738, doi:10.1016/j.tree.2016.07.007 (2016).

127 Jamil, T., Ozinga, W. A., Kleyer, M. & ter Braak, C. J. F. Selecting traits that explain species–environment relationships: a generalized linear mixed model approach. *Journal of Vegetation Science* **24**, 988-1000, doi:10.1111/j.1654-1103.2012.12036.x (2013).

128 Abrego, N., Norberg, A. & Ovaskainen, O. Measuring and predicting the influence of traits on the assembly processes of wood-inhabiting fungi. *Journal of Ecology*, doi:10.1111/1365-2745.12722 (2017).

129 Ferrier, S., Manion, G., Elith, J. & Richardson, K. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions* **13**, 252-264, doi:10.1111/j.1472-4642.2007.00341.x (2007).

130 Rose, P. M., Kennard, M. J., Sheldon, F. & Moffatt, D. B. A data-driven method for selecting candidate reference sites for stream bioassessment programs using generalised dissimilarity models. *Marine and Freshwater Research* **67**, 440-454 (2015).

131 Woolley, S. N. C., Foster, S. D., O'Hara, T. D., Wintle, B. A. & Dunstan, P. K. Characterising uncertainty in Generalised Dissimilarity Models. *Methods in Ecology and Evolution*, doi:10.1111/2041-210X.12710 (2017).

132 Rosauer, D., F. *et al.* Phylogenetic generalised dissimilarity modelling: a new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. *Ecography* **37**, 21-32, doi:10.1111/j.1600-0587.2013.00466.x (2014).

133 Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters* **18**, 1-16, doi:10.1111/ele.12376 (2015).

134 Mokany, K. *et al.* Integrating modelling of biodiversity composition and ecosystem function. *Oikos* **125**, 10-19, doi:10.1111/oik.02792 (2016).

135 Willis, K. J. *et al.* Determining the ecological value of landscapes beyond protected areas. *Biological Conservation* **147**, 3-12, doi:10.1016/j.biocon.2011.11.001 (2012).

136 Willis, K. J. *et al.* Remote assessment of locally important ecological features across landscapes: how representative of reality? *Ecological Applications* **25**, 1290-1302, doi:10.1890/14-1431.1 (2015).

137 Leitão, P. J. *et al.* Mapping beta diversity from space: Sparse Generalised Dissimilarity Modelling (SGDM) for analysing high-dimensional data. *Methods in Ecology and Evolution* **6**, 764-771, doi:10.1111/2041-210X.12378 (2015).

138 Reside, A. E. *et al.* Climate change refugia for terrestrial biodiversity: Defining areas that promote species persistence and ecosystem resilience in the face of global climate change. 216 (National Climate Change Adaptation Research Facility, Gold Coast, 2013).

139 Allnutt, T. F. *et al.* A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. *Conservation Letters* **1**, 173-181, doi:10.1111/j.1755-263X.2008.00027.x (2008).

140 Mokany, K., Harwood, T., Overton, J., Barker, G. & Ferrier, S. Combining α - and β - diversity models to fill gaps in our knowledge of biodiversity. *Ecology letters* **14**, 1043-1051, doi:10.1111/j.1461-0248.2011.01675.x (2011).

141 Mokany, K., Prasad, S. & Westcott, D. A. Impacts of climate change and management responses in tropical forests depend on complex frugivore-mediated seed dispersal. *Global Ecology and Biogeography* **24**, 685-694, doi:10.1111/geb.12292 (2015).

142 Schuldt, A. *et al.* Multitrophic diversity in a biodiverse forest is highly nonlinear across spatial scales. *Nature Communications* **6**, 10169, doi:10.1038/ncomms10169 (2015).

143 Wisz, M. S. *et al.* The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* **88**, 15-30, doi:10.1111/j.1469-185X.2012.00235.x (2013).

144 D'Amen, M., Rahbek, C., Zimmermann, N. E. & Guisan, A. Spatial predictions at the community level: from current approaches to future frameworks. *Biological Reviews* **92**, 169-187, doi:10.1111/brv.12222 (2017).

145 Urban, M. C. *et al.* Improving the forecast for biodiversity under climate change. *Science* **353**, doi:10.1126/science.aad8466 (2016).