This is a repository copy of *Connecting Earth Observation to High-Throughput Biodiversity Data*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/118918/

Version: Accepted Version

Article:

White, Piran Crawfurd Limond orcid.org/0000-0002-7496-5775 and Yu, Douglas (2017) Connecting Earth Observation to High-Throughput Biodiversity Data. Nature Ecology and Evolution. pp. 1-9.

https://doi.org/10.1038/s41559-017-0176

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

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, ǂ]

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).

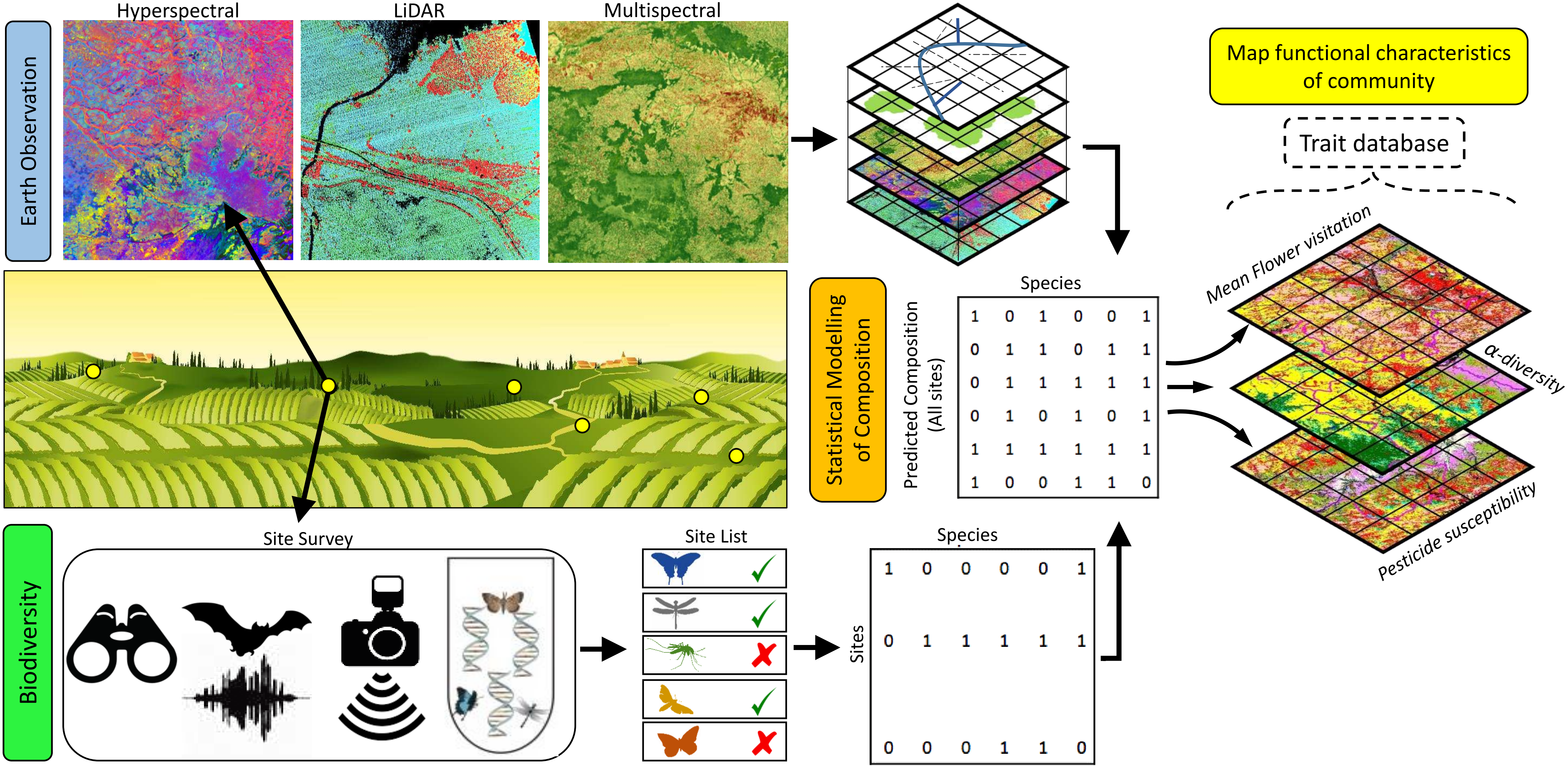
Figure legends

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/).

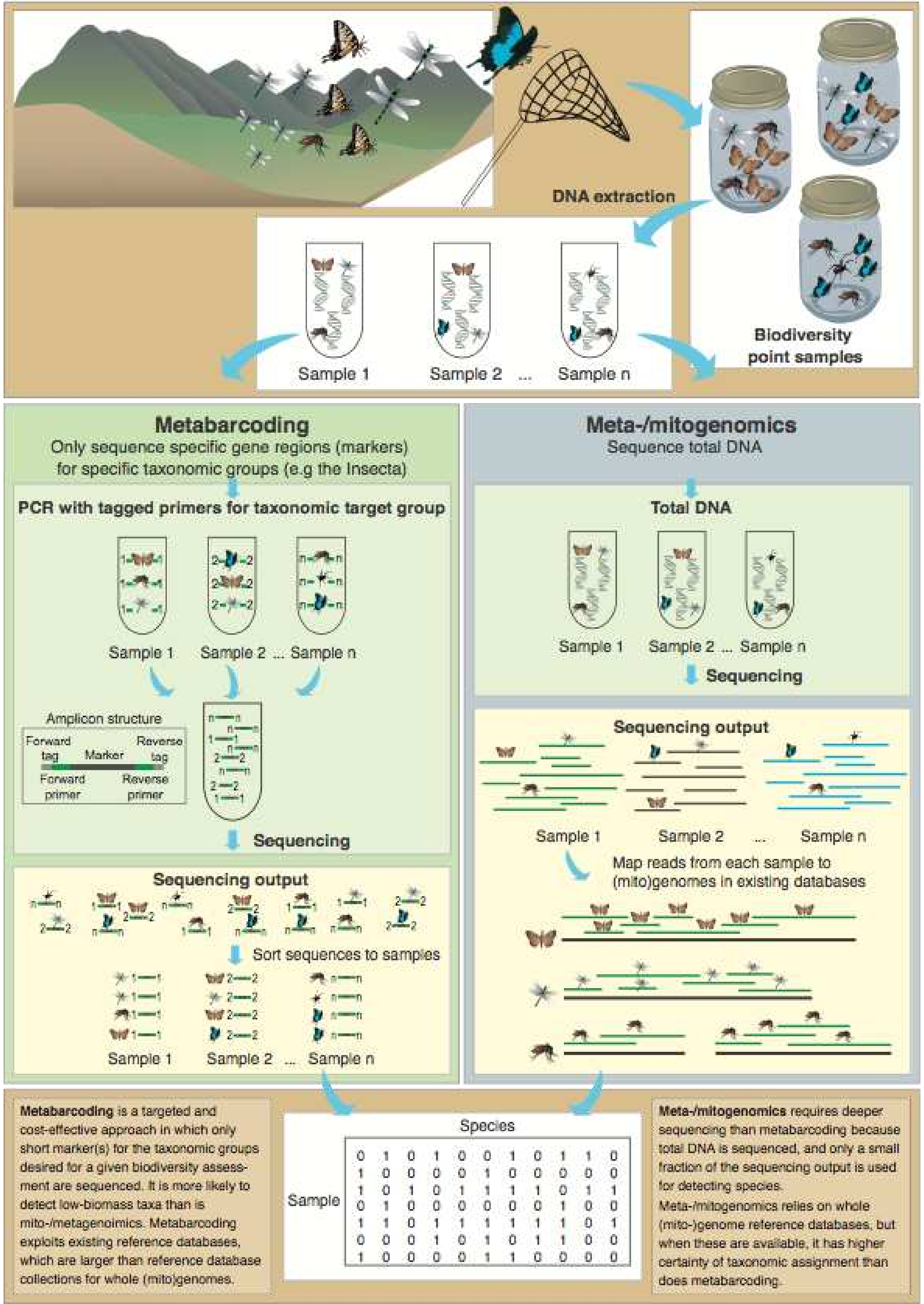
Figure 3. Three statistical pathways to map community composition and summary metrics from the combination of biodiversity point samples and continuous Earth Observation (EO) maps. Local diversity – α; species turnover – β; and regional diversity – γ. For clarity, the figure only considers models for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



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

	Species					
Sites	1	0	0	0	0	1
	0	1	1	1	1	1
	0	0	0	1	1	0

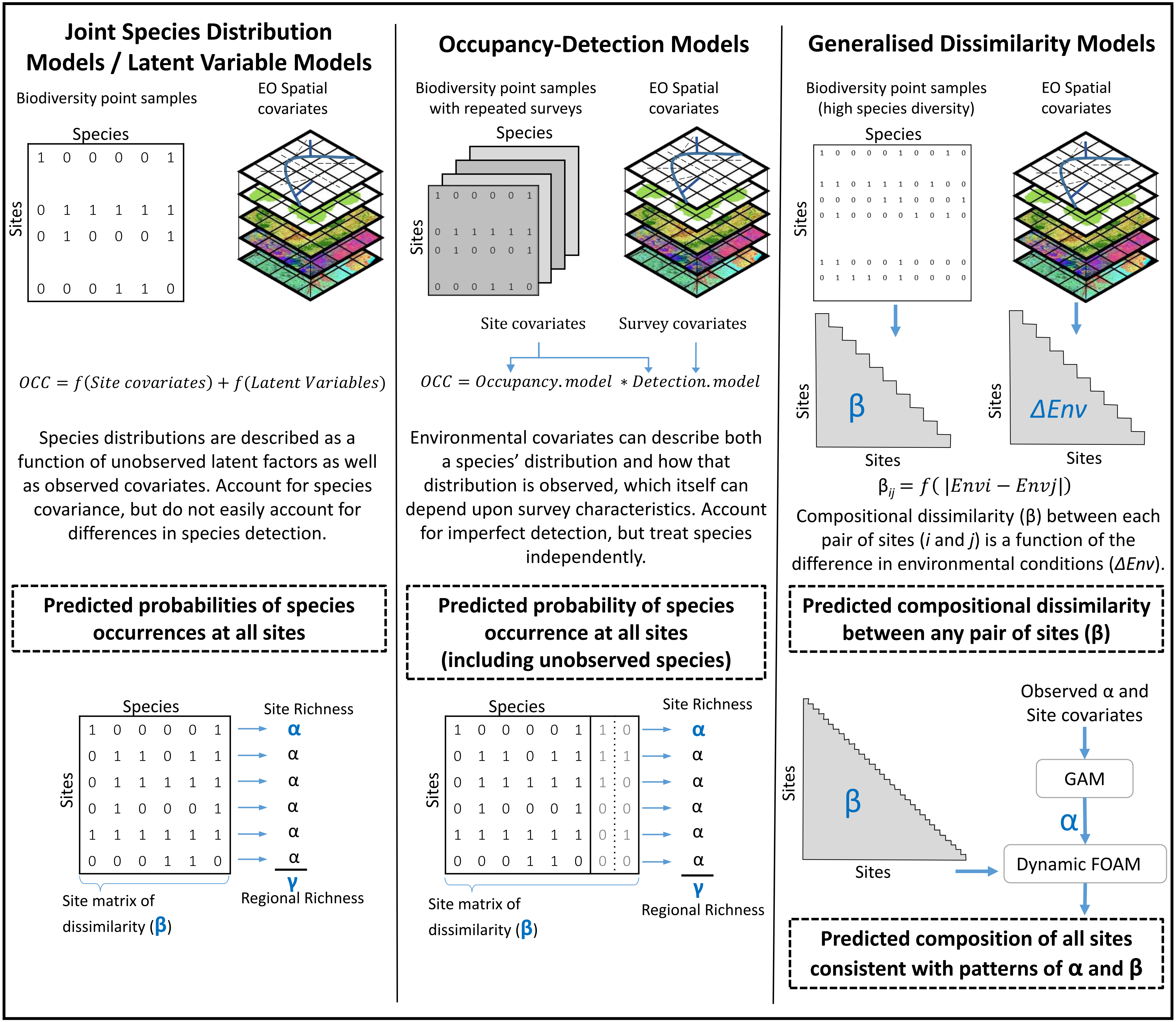


Metabarcoding is a targeted and cost-effective approach in which only short marker(s) for the taxonomic groups desired for a given biodiversity assessment are sequenced. It is more likely to detect low-biomass taxa than is mito-/metagenoimics. Metabarcoding exploits existing reference databases, which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

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

Site matrix of dissimilarity (β)

Regional Richness

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

Site matrix of dissimilarity (β)

Regional Richness

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

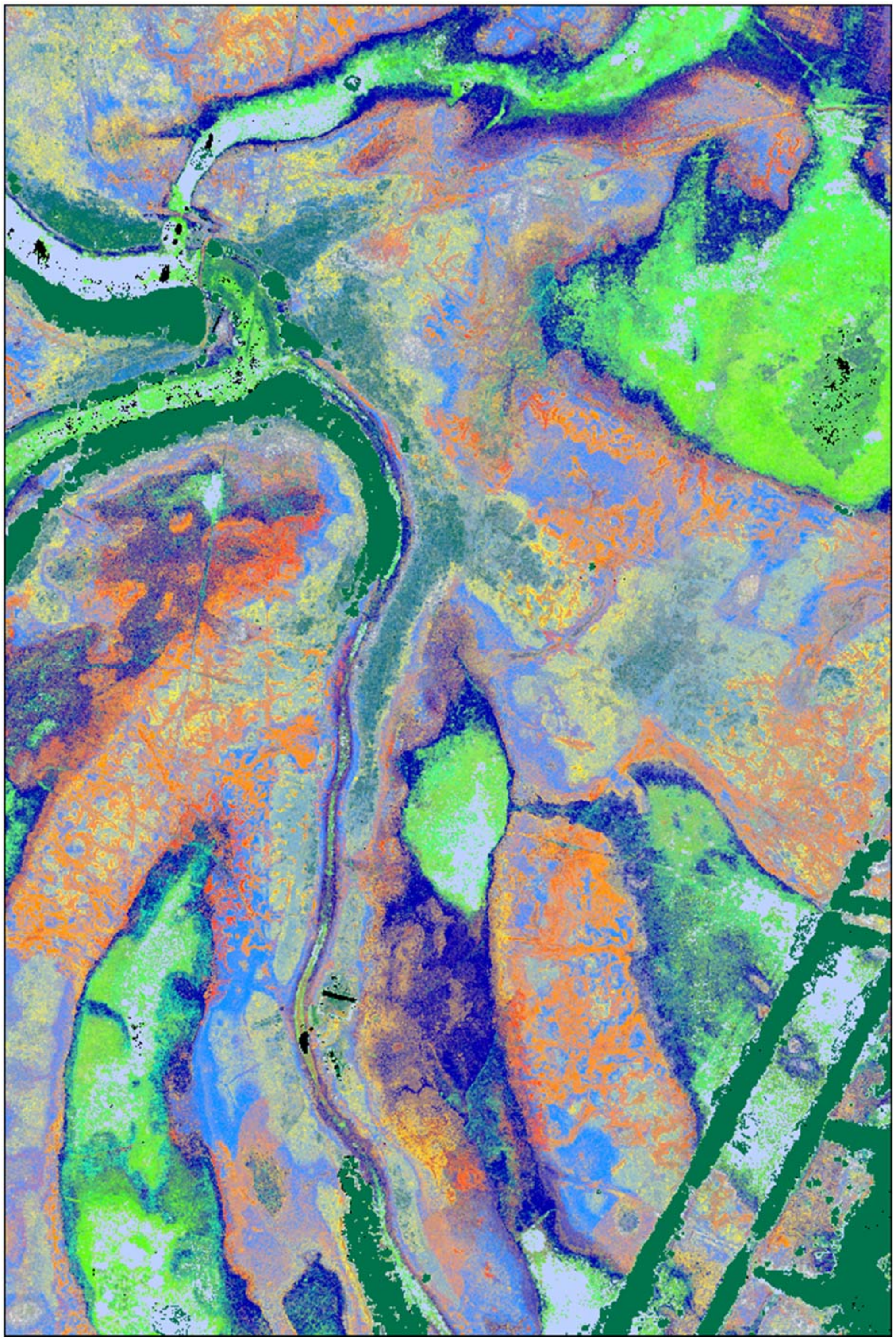
1 1 0 0 0 1 0 0 0 0
0 1 1 1 0 1 0 0 0 0

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

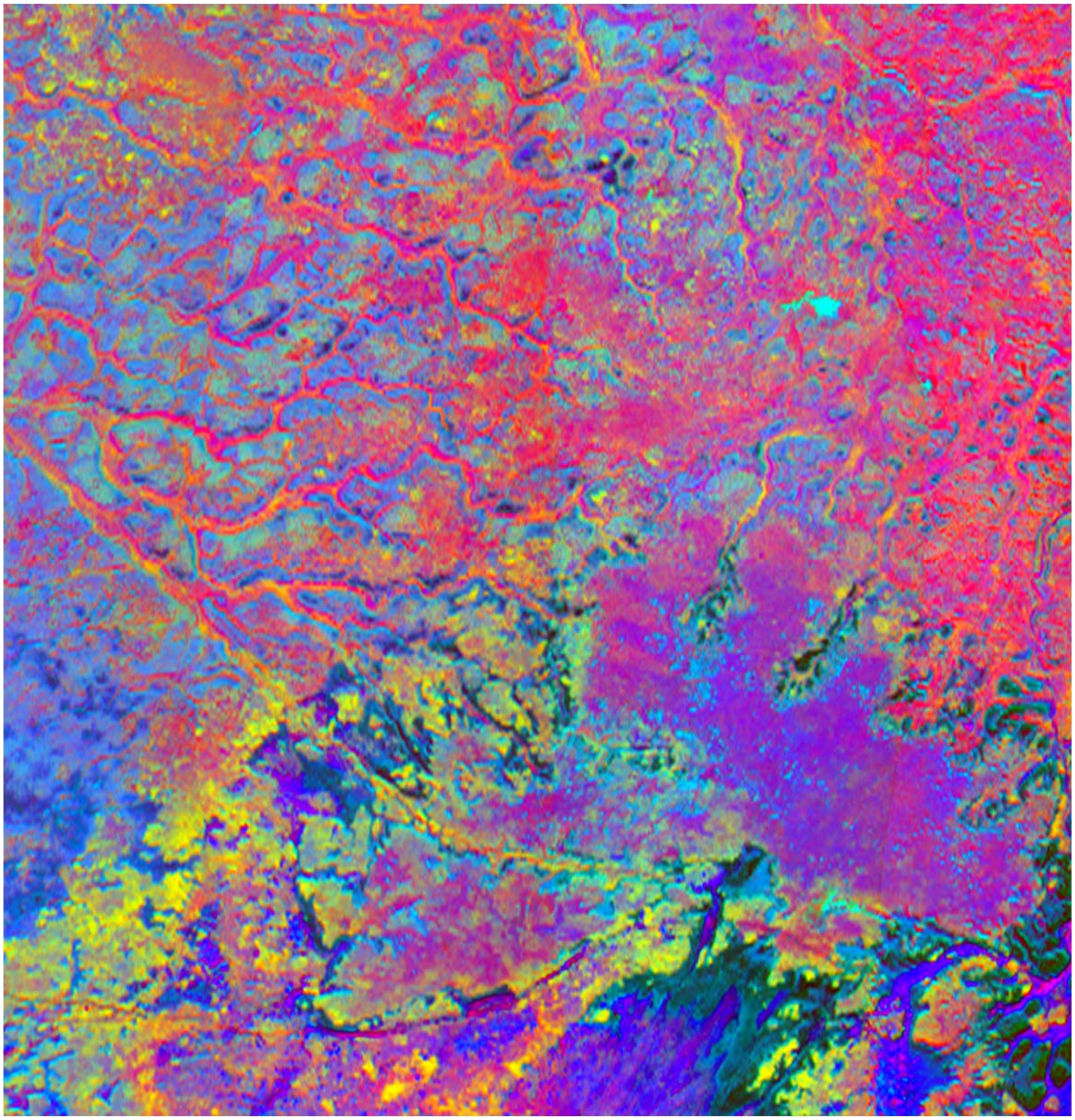
Compositional dissimilarity (β) between each pair of sites (*i* and *j*) is a function of the difference in environmental conditions (*ΔEnv*).

Predicted compositional dissimilarity between any pair of sites (β)

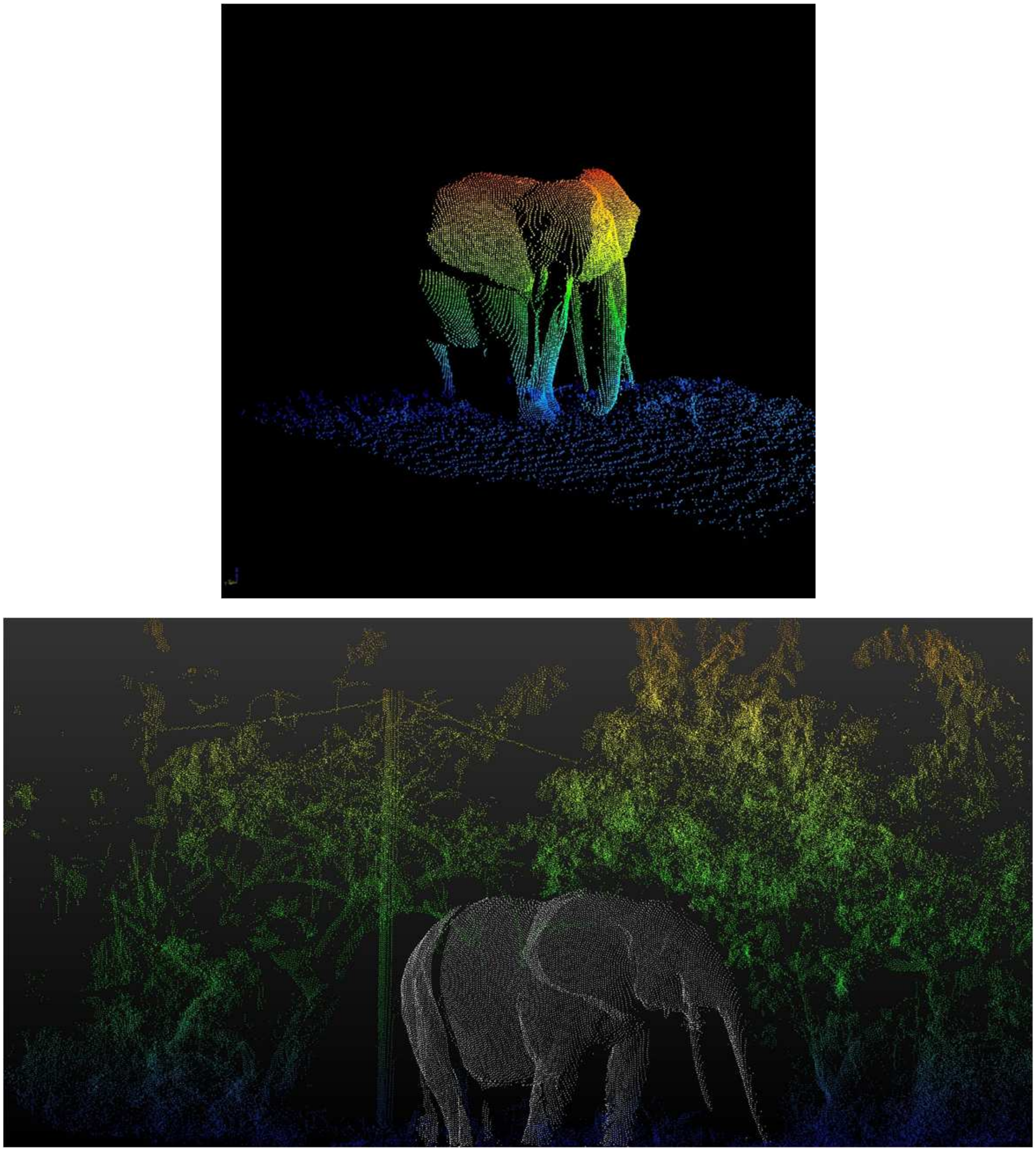
Predicted composition of all sites consistent with patterns of α and β



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).