Supporting global biodiversity assessment through high-resolution macroecological modelling: Methodological underpinnings of the BILBI framework.

Andrew J Hoskins[1,a,*], Thomas D Harwood[1], Chris Ware[1], Kristen J Williams[1], Justin J Perry[2], Noboru Ota[3], Jim R Croft[4], David K Yeates[5], Walter Jetz[6,7], Maciej Golebiewski[8], Andy Purvis[9], Simon Ferrier[1]

[1] CSIRO Land and Water, Canberra, ACT, Australia

[2] CSIRO Land and Water, ATSIP, Aitkenvale, Queensland, Australia

[3] CSIRO Agriculture and Food, Wembley WA, Australia

[4] Australian National Botanic Gardens, Canberra, ACT Australia

[5] Australian National Insect Collection, CSIRO National Research Collections Australia, ACT, Australia

[6] Yale University, Department of Ecology and Evolutionary Biology, New Haven, CT, USA

[7] Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire, UK

[8] CSIRO High Performance Computing and Communications Centre, Melbourne, Australia

[9] Natural History Museum, London, UK

[a] Current address: CSIRO Health and Biosecurity, ATSIP, Aitkenvale, Queensland, Australia.

[*] Corresponding author: Andrew.Hoskins@csiro.au

Abstract

Efficient planning for biodiversity-related goals requires the ability to measure biodiversity's present state, observe recent change and project potential future outcomes from predicted anthropogenic change scenarios. Indicators of biodiversity change are most often derived by intersecting changes in habitat loss or degradation with underlying patterns in biological diversity. However, the current generation of these indicators are usually derived at resolutions far coarser than the spatial grain that key ecological and anthropogenic processes they are describing interact. Concurrently, current indicators trend to focusing on data rich biological groups, ignoring the other parts of the tree of life where data are less complete. Here we introduce the methodological underpinnings to a new capability in global biodiversity assessment that is based around macroecological modelling of underlying patterns in biodiversity. Through integrating advances in macroecological modelling, biodiversity informatics, remote sensing and high-performance computing, our modelling infrastructure assesses the spatio-temporal change in biodiversity across the entire terrestrial surface of the planet at a resolution of approximately 1 km. We then use this to follow two different analysis pathways. The first pathway generates indicators of biodiversity change by overlaying modelled biodiversity pattern with observed patterns in land-use change to observe the changing status and trends in ecological communities. Our second pathway replaces observed changes with projected changes in pressures and responses into the future. This enables the translation of alternative global change scenarios into the expected consequences for future biodiversity persistence. This approach allows a more robust assessment of the status, trends and possible futures for as broad a spread of biological diversity as possible.

INTRODUCTION

Continued growth in human populations around the world is intensifying demands on our natural environment. Coupled with the effects of anthropogenic climate change, the potential for large-scale modification and loss of our planet's remaining biological diversity seems ever more likely (Pereira *et al.*, 2010). To combat this ongoing decline, governments have agreed to multi-lateral policy goals which aim to limit, reduce or halt biodiversity loss and environmental degradation. The Convention on Biological Diversity (CBD) Strategic Plan for Biodiversity 2011-2020 and the associated Aichi Biodiversity Targets are one such policy framework that sets near-future targets across five strategic goals addressing ultimate drivers of biodiversity loss, proximate pressures, management responses, benefits to people, and implementation challenges (CBD, 2010). More recently the Sustainable Development Goals (SDGs) adopted by the United Nations promote a healthy and sustainable future both for humans and for our environment, including all "life on land" and "life below water" (UN, 2015), while the latest multi-lateral agreement to limit anthropogenic climate change, ratified in Paris in 2015, includes statements to limit the loss of natural habitat (through deforestation) with indirect consequences for biodiversity (Citroen *et al.*, 2016).

Efficient planning of actions to achieve biodiversity-related goals and targets under these policy processes, and effective tracking of progress towards this achievement, requires the ability to measure the present state of biodiversity, detect trends of recent change, and project the potential future state of biodiversity expected under alternative policy options in a globally consistent way. Unfortunately our ability to report or project indicators of change for many aspects of biodiversity is still limited by an inability to observe or infer changes in ecological communities directly from currently available global datasets (Ferrier, 2011). Indicators of change employed in biodiversity assessments are most often derived by intersecting observed or projected changes in the distribution of habitat loss and degradation, or of protected areas, with underlying patterns in the distribution of biodiversity (e.g. Tittensor *et al.*, 2014; Butchart *et al.*, 2015)

Two sources of global data on terrestrial biodiversity patterns have been used most commonly in the derivation of protected-area and habitat indicators. The first of these is the World Wildlife Fund's mapping of 867 terrestrial ecoregions, defined as "relatively large units of land containing a distinct assemblage of natural communities and species, with boundaries that approximate the original extent of natural communities prior to major land-use change" (Olson *et al.*, 2001). Ecoregions have long provided a convenient and well-respected foundation for assessing changing patterns of protected-area coverage and habitat transformation around the world (e.g. Watson *et al.*, 2016). However, as indicated by the above definition, ecoregions are typically delineated at a much coarser resolution than the spatial grain of key ecological processes shaping both land-use and biological distributions at the landscape scale (Londoño-Murcia *et al.*, 2010; Calderón-Patrón *et al.*, 2016).

Using ecoregions as fundamental spatial units for assessing impacts of protected-area coverage and habitat transformation on biodiversity assumes that all biological elements (e.g. species) within an ecoregion will be equally affected by these activities. Yet, in reality, fine-scaled spatial heterogeneity in abiotic environmental attributes (e.g. terrain, soils, climate) within an ecoregion will tend to bias human uses to particular parts of the region (e.g. a greater likelihood of agriculture in flatter, more fertile environments) (Fig. 1a). Since these same environmental attributes shape natural distributions of species at landscape scale (Fig. 1b), impacts of any given land-use change within an ecoregion will tend to be biased towards a subset of the species occurring within that region (Fig. 1c). This means that protected-area or habitat indicators derived using ecoregions as the fundamental units of analysis risk under- or over-estimating implications of protection or habitat

transformation for biodiversity contained within these regions (Ferrier *et al.*, 2004; Londoño-Murcia *et al.*, 2010).

The other major source of global data on biodiversity patterns commonly used for indicator derivation – i.e. extent-of-occurrence range maps for terrestrial vertebrate species (e.g. Jenkins *et al.*, 2013) – presents similar spatial-resolution challenges. As for ecoregions, this data source has, over recent years, enabled the derivation of a wide variety of indicators, and has also underpinned numerous macroecological analyses of global biodiversity patterns. However the relatively coarse resolution of most range maps, and the reality that many species occupy only those parts of their overall range offering suitable environmental conditions, has led some workers to suggest that these data should not be employed at a grid resolution finer than 1 degree, or approximately 100km x 100km near the equator (Hurlbert & Jetz, 2007). This again is a resolution far coarser than the spatial grain of key ecological processes shaping land-use and biological distributions at landscape scale.

Species distribution modelling (SDM) provides one widely used means of refining the resolution of mapped species distributions, by using fine-resolution environmental surfaces to characterise and spatially project a species' niche space (Elith & Leathwick, 2009). This can be achieved either by using known occurrence records to fit a correlative model predicting occurrence of a given species as a mathematical function of multiple environmental variables, or through deductive modelling in which occurrence is predicted using simple rule-based descriptions of environmental suitability derived from expert knowledge (Ferrier, 2002). Distributions predicted using SDM can be used either directly in assessments, or combined with mapped species ranges, where available (e.g. for vertebrates), thereby providing refined mapping of the expected distribution of each species within its known range (Merow *et al.*, 2017). However, regardless of the precise SDM technique employed, application of this general approach is restricted to species for which either there is a sufficient number of occurrence records available to develop a correlative model, or there is sufficient expert knowledge of the species' habitat requirements to develop a deductive model. This capacity is therefore limited to a subset of species which is biased both taxonomically and geographically, with some regions of the world lacking adequate data to generate reliable SDMs even for better-known biological groups such as vertebrates, let alone for invertebrates and plants (Meyer *et al.*, 2015).

Here we adopt an alternative, yet highly complementary, approach to integrating species-occurrence records with fine-scaled environmental surfaces to refine the spatial resolution with which patterns in the distribution of biodiversity can be mapped across our planet. Rather than attempting to model distributions of individual species, this approach instead focuses on modelling, and thereby mapping, collective properties of biodiversity as a correlative function of environmental predictors. Macroecological modelling of spatial variation in alpha diversity, particularly of variation in local species richness, has a relatively long history of application in ecology and conservation biology (e.g. Francis & Currie, 1998). However, with increasing awareness that the total (gamma) diversity encompassed by any set of areas (e.g. in a conservation reserve system) will typically depend more on the extent to which these areas complement one another in terms of species composition, than it does on the richness of individual areas, macroecological modelling is now placing greater emphasis on modelling patterns of beta diversity in addition to those of alpha diversity (Ferrier & Guisan, 2006; D'Amen *et al.*, 2017).

We here introduce a new capability for global biodiversity assessment – BILBI (the Biogeographic Infrastructure for Large-scaled Biodiversity Indicators) – underpinned by macroecological modelling of collective properties of biodiversity. The initial implementation of this infrastructure relies strongly on modelling of beta diversity patterns using an extension of one particular technique – generalised dissimilarity modelling (GDM; Ferrier *et al.*, 2007) – applied to readily available biological

and environmental datasets. The overall framework is, however, designed to be sufficiently generic and flexible to allow incremental refinement and addition of modelling techniques and datasets into the future. This capability is also intended to complement, rather than compete with, other approaches to global biodiversity assessment, including those focussed on individual species (e.g. Jetz *et al.*, 2012). Species-level approaches will always play a vital role in biodiversity assessment for better-known biological groups, and especially for species of particular conservation concern within these groups. However the approach described here has potential to add significant value to such species-based assessments by: 1) allowing more effective use of data for highly-diverse biological groups, containing large numbers of species but with few records per species; and 2) enabling robust extrapolation of expected patterns across poorly-sampled regions, even where the particular species occurring in these regions are unknown or unsurveyed.

METHODS

General approach

The BILBI modelling framework (Fig. 2) integrates advances in macroecological modelling, biodiversity informatics, remote sensing and high-performance computing to assess spatial-temporal change in biodiversity at 30-arcsecond (approximately 1km) grid resolution across the entire terrestrial surface of the planet. Best-available data on observed occurrences of species within defined biological groups (e.g. all vascular plants) are used to fit correlative models describing patterns in the distribution of biodiversity as a function of fine-scaled spatial variation in climate, terrain and soils, within major habitat types (biomes) and biogeographic realms. These patterns are mapped as spatially-complete gridded surfaces by interpolating and, where necessary, extrapolating predictions from the fitted models. The resulting surfaces describe patterns in the spatial distribution of biodiversity which would be expected in the absence of anthropogenic habitat transformation. These modelled patterns then serve as the foundation for two subsequent pathways of analysis in the BILBI framework (Fig. 2).

In the first pathway these patterns of biodiversity distribution are overlayed with observed changes in pressures (direct drivers) – particularly changes in habitat condition resulting from land-use change – or in management responses, such as the establishment of protected areas, to generate indicators of biodiversity change (e.g. for reporting progress towards the CBD's Aichi Targets). In the second pathway, observed changes are replaced by projected changes in pressures and responses into the future. This enables application of BILBI in translating alternative scenarios of global change, and associated policy or management options, into expected consequences for the future persistence of biodiversity. In assessing such scenarios the BILBI framework allows consideration both of impacts mediated by changes in habitat condition, resulting for example from projected land-use change, and of potential impacts of climate change on community composition. The latter is predicted through space-for-time substitution of climate covariates in BILBI's correlative models of spatial biodiversity distribution.

In the remainder of this paper we describe our initial implementation of the BILBI framework, focusing primarily on the foundational modelling of spatial patterns in the global distribution of terrestrial biodiversity. The two analytical pathways flowing from this foundation – relating to indicator generation, and scenario analysis respectively (Fig. 2) – will be addressed in detail in subsequent papers. As alluded to above, our modelling of biodiversity patterns in the initial implementation of BILBI has focused on describing and predicting spatial turnover in species composition – i.e. patterns of beta diversity. However, our longer-term intent is to extend this

approach to accommodate joint modelling of both alpha and beta diversity; and to integrate next-generation techniques for achieving this as they become operational.

Modelling compositional turnover using presence-only data

GDM is a nonlinear regression technique for modelling the turnover in species composition expected between two sites as a function of environmental differences between, and geographical separation of, these sites. This technique accommodates two types of nonlinearity commonly encountered in large-scaled analyses of compositional turnover. The curvilinear relationship between increasing environmental or geographical distance, and observed compositional dissimilarity, between sites is addressed through the use of appropriate link functions in a generalised linear modelling framework. Variation in the rate of compositional turnover at different positions along environmental gradients is addressed by transforming these gradients using smooth monotonic functions fitted to the training data (Ferrier *et al.*, 2007). The response variable in a standard GDM model is typically a measure of between-site compositional dissimilarity, calculated from lists of species observed at each of the two sites, using indices such as Sørensen or βsim (e.g. Jones *et al.*, 2013; König *et al.*, 2017). However, one of the biggest challenges in applying GDM globally has been that a large proportion of available species-occurrence data are presence-only rather than presence-absence in nature. Most occurrence records accessible through major data infrastructures, such as the Global Biodiversity Information Facility (GBIF), have been generated through geo-referencing of specimens from natural-history collections, or from relatively opportunistic field observations of individual species, rather than from planned inventories systematically recording all species present at a given site (Isaac & Pocock, 2015). Such data are not well suited to estimating compositional dissimilarity between sites, particularly in areas with lower sampling effort. This is because estimates of compositional dissimilarity will be inflated, to a varying yet unknown extent, by false absences of species at each of the sites concerned (Beck *et al.*, 2013).

In implementing the BILBI framework we have addressed this problem by modifying GDM to work with a binary response variable, defined in terms of matches versus mismatches in species identity, for pairs of individual species observations (where a "species observation" is the recorded presence of a particular species at a particular site). The probability that a species randomly drawn from site *i* has the same identity as a species randomly drawn from site *j* is expected to be a function of both the total number of species actually occurring at each of the two sites (alpha diversity) and the number of species at each site which are unique to that site, because they do not occur at the other site (beta diversity), following the expression:

$$p_{i,j} = \sum_{k=1}^{n} \left(\frac{[s_{k,i}]}{\alpha_{k,i}} \times \frac{[s_{k,j}]}{\alpha_{k,j}} \right)$$

$$[s] = \begin{cases} 1, & s \text{ is found at the site of interest;} \\ 0, & s \text{ is not found at the site of interest.} \end{cases}$$

where *s* is the identity of an individual species belonging to the combined list of species occurring at either one, or both, of the sites (n) and α is the number of species found at a particular site. As the quantity being summed reduces to 0 when *s* is not shared by both sites, this equation can be simplified to to $p_{i,j} = \left(1/[\alpha_i\ \alpha_j]\right)c$, where c is the number of species shared between the two sites.

Using this understanding, we fit a modified form of GDM in which the response variable is a binary description of the match (0) or mismatch (1) in species identity of a randomly drawn pair of species

observations from two sites. This observation-pair GDM (*obs-pair*GDM) is then modelled as a non-linear combination of environmental covariate data following

$$logit(1 - p_{i,j}) = \beta_0 + \sum_{l=1}^{m} \left| f_l(x_{l,i}) - f_l(x_{l,k}) \right| + \varepsilon$$

where f represents a combination of I-spline basis functions (Ramsay, 1988). We can now model $p_{i,j}$ as a function of the environmental separation between two sites in the same way as a standard GDM, by ensuring $\left| f_l(x_{l,i}) - f_l(x_{l,k}) \right|$ represent the separation of a pair of sites in the spline transformed environmental space. This is not yet an estimate of the turnover in species composition between two sites. As we showed above, $p_{i,j}$ is a function not only of compositional turnover, but also of the richness of species at the two sites concerned. However, if we can estimate species richness then we can decompose this value into an estimation of compositional turnover between any pair of sites (β_{ij}) using

$$\beta_{ij} = 1 - \frac{p_{ij}}{p_0}$$

*w*here p_0 represents an estimate of the probability that a randomly drawn pair of species from a pair of identical sites (i.e. compositional turnover equals 0) are the same. To enable fitting we estimate p_0 as the intercept of our model – the point where environmental separation between sites is 0 and thus the sites are treated as the same.

By fitting our models to pairs of individual species observations this method avoids the biases that can result from modelling community data where the inventory of species at sites is incomplete. We just need to satisfy the assumption that the particular species recorded as present at a given site constitute a random sample drawn from all species actually occurring at that site. In its current form the method also assumes that species richness remains reasonably constant across the region of interest – i.e. that the number of species actually occurring at individual sites (1km cells in this study) does not vary markedly across the region – and therefore that the effect of alpha diversity on the response being modelled is accounted for by the model's intercept. As we describe in the next section, we have taken considerable care to minimise violation of this assumption by fitting separate models for different biomes and biogeographic realms. Our team is also currently developing an extension of the above approach which relaxes this assumption, and thereby explicitly models $p_{i,j}$ as a function of variation in both alpha and beta diversity. Preliminary testing suggests that that this approach holds considerable promise as a means of simultaneously modelling patterns of both species richness and compositional turnover from presence-only data.

Initial global implementation – model fitting

A suite of models, describing compositional-turnover for species in three biological groups (invertebrates, vascular plants and vertebrates) across the terrestrial surface of the globe (above 60˚S, i.e. excluding Antarctica) was generated using this *obs-pair*GDM technique (e.g. Fig. 3a-d). Models were fitted by pairing the best-available climate, terrain and soil surfaces with data from GBIF and Map of Life (MoL; Jetz *et al.*, 2012) to produce global estimates of compositional-turnover at 30-second resolution (described in more detail below).

Models were developed within the WWF nested realm, biome and ecoregion framework, with separate models fitted to each of the 61 unique realm/biome or bio-realm intersections. This allowed consideration of major biogeographic discontinuities between realms and potential variation in the response of different species assemblages between biomes. However, biomes were

as far as possible not treated as discrete units, so models drew on biological and environmental data from both the core biome and adjacent biomes within the realm, and from adjacent ecoregions in neighbouring realms where the realm boundary was considered porous to the movement of species (e.g. the Nearctic/Neotropics divide in Florida). Models were then fitted to a combination of pairs of observations taken 50% from sites exclusively within the target bio-realm and 50% from pairs of sites taken between the target bio-realm and the buffering regions

Our models used three spline basis functions with knots placed at the 0, 0.5 and 1 quantiles of the environmental space contained within each target bio-realm. Where the environmental envelope of samples taken from buffering regions lay outside the envelope for the target biome, additional knots were added at the outer limits, resulting in 3, 4 or 5 knots depending on the structure of the data. This allowed us to maximise the description of spatial pattern within our primary analysis area (the target bio-realm) while also incorporating potential additional spatial pattern arising from these neighbouring regions.

Geographic distance is generally a very good descriptor of spatial turnover in species composition, but both describes the effects of biogeographical separation and acts as a proxy descriptor for the effects of environmental gradient. In this latter context, it is important to avoid geographic distance overwhelming the effects of more direct descriptors within the models. We therefore fitted our models in a two-staged procedure; firstly fitting to the *a priori* selected set of environmental predictor variables, and then refitting the models with an offset, allowing geographic distance to describe only variation not already described by the environmental covariates. Geographic distance was fitted using a linear function, rather than the complex splined functions used for environmental covariate data.

Initial global implementation – biological inputs

Biological data on species observations was obtained by downloading the complete GBIF database for vascular plants, invertebrates and reptiles. These data were paired with the entire MoL database for birds, amphibians and mammals for final analyses. All data were then filtered to remove erroneous points falling outside of our land mask (e.g points falling in the ocean) with records found to occur ≤ 1 km from the coastline, and belonging to terrestrial species, moved to fall within the nearest terrestrial cell. Following this, terrestrial species where resources are primarily obtained from the marine environment (e.g. procellariform seabirds and pinniped) were removed from the dataset. After consultation with our taxonomic experts for vascular plants (JR Croft) and invertebrates (D Yeates) we reduced the families within the invertebrate group to a subset of families where species are predominantly terrestrial (or in a terrestrial phase of their life cycle when sampled) and the communities of practice surrounding the recording of these species are similar (see list in supplementary materials).

The final set of species records was then aggregated into individual 30-arc second resolution sites concordant with the spatial grain of the environmental surfaces being used in the modelling. Data were then filtered to contain only unique species records per unique spatial site. Filtering then occurred to remove sites where an extreme number of species (in relation to surrounding sites), likely representing the location of biological collections rather than actual species locations. The final pool of data resulted approximately 300,000 individual species and 100,000,000 individual records being included in the modelling procedure (Table 1). We sampled a maximum of 1.5 million samples from each of the unique realm-biome datasets. Samples further stratified so that within the broad biological groups, families/taxon where there existed different communities of practice of the collection of these data, and thus the potential for bias/skew in the distribution of these data, no

pair-wise samples were taken between these groups (See Table S1 for groupings). These 1.5 million samples were evenly sampled from each of these separate lower level biological groupings, however where the desired number of samples could not be drawn from the data, the remaining groups were not restricted to only this lesser number of samples. Additionally, to avoid a skewed influence of a small number of species observations within any model, each single species record was only used a maximum of 1 times during the development of sample pairs.

Initial global implementation – environmental and land-use inputs

Environmental covariate data were selected *a priori* from the suite of possible bioclimatic, terrain and soil predictors. Individual model selection of covariate sets was not performed as this would reduce the comparability of results between models and limit our ability to generate continuous surfaces of biological composition across large areas. This *a priori* selection, drew on significant variables from past GDM models fitted by the team across a range of scales and biota, and an attempt to capture ecologically limiting factors across the world. Additional criteria were that the layers were freely available with consistency of global coverage.

A set of 15 standard candidate environmental variables were prepared as continuous global surfaces (Table 2). The grids comprise 5 soil variables (bare ground, bulk density, clay, pH, silt), 2 terrain variables (topographic roughness index, topographic wetness index) and 8 climate variables (annual precipitation, annual minimum temperature, annual maximum temperature, maximum monthly diurnal temperature range, annual actual evaporation, potential evaporation of driest month, maximum and minimum monthly water deficit). For climate variables it was important that they could be consistently projected through time, and so the WorldClim (Hijmans *et al.*, 2005) elevation adjusted data set was chosen over a remotely sensed product. Adjustment for radiative shading based on the GMTED2010 DEM (Danielson & Dean, 2011), and derivation of evaporation, water balance and summary statistics was carried out as in (Reside *et al.*, 2013a). All grids were aligned to the WorldClim 30s land extent, with water bodies (defined as the Global Lakes and Wetlands Database v3 Lakes and Reservoirs: Lehner & Doll, 2004) masked out. Where necessary for the soil and terrain layers, minor information gaps were filled using a combination of extrapolation and appropriate values drawn from the literature. All methods developed and used to create these layers were designed to the applicable to both present day bioclimatic data and future scenarios generated from General Circulation Models (GCM).

As previously mentioned, the other method that impacts can be input into BILBI is through the use of condition layers that scale the local impacts of land-use and habitat protection/conversion. We have aggregated multiple different layers for use, depending on the desired analysis. The World Database on Protected Areas (WDPA) that is curated by the World Conservation Monitoring Centre (UNEP & WCMC, 2016) is ingested as a condition layer when running analyses to assess the effectiveness of protected area coverage. Loss of forest detected by the global forest loss dataset (Hansen *et al.*, 2013) is used to analyse the biodiversity impacts of forest habitat loss. We have also developed methods and datasets for global land-use at 30-arc second resolution by downscaling coarse grained land-use data (Hurtt 2011). By teaming with the PREDICTS project (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems; Hudson *et al.*, 2014) we can convert these land-use layers into estimates of the loss of local species diversity (species richness; e.g. Gray *et al.*, 2016; Newbold *et al.*, 2016; Palma *et al.*, 2017) along a 0 – 1 scale, which then allows us to infer regional land-use impacts (though loss of compositional similarity) to biodiversity using BILBI.

Figures and Tables

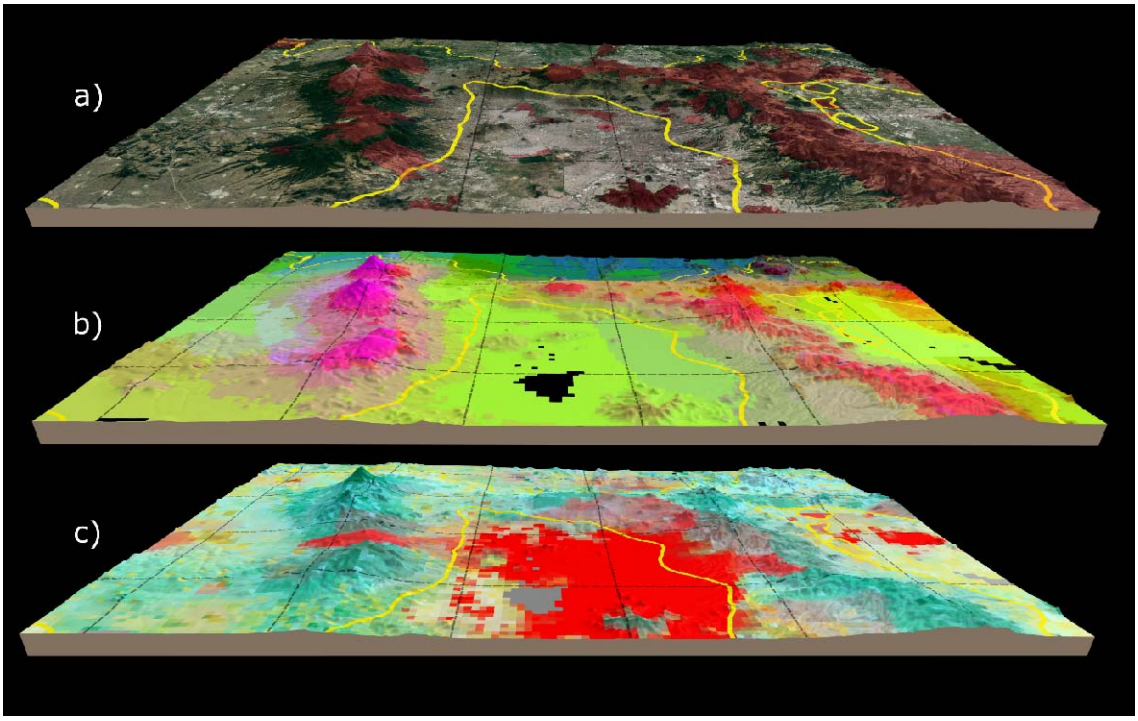

Figure 1: Depiction of the problem of scale when making biodiversity assessments. (a) shows a true colour satellite overlay of the region surrounding Mexico city with typical resolutions for global analyses – a 30 km^2 grid (black) and ecoregional boundaries (yellow) – and protected areas overlayed in red. Note the geographic/topographic biases in placement of both protected areas and urban development. (b) shows the same region but the different ecological communities (defined by similarity between communities in each 1 km^2 pixel) where similar colours represent similar communities. (c) shows land-use from Hoskins *et al.* (2016) where red depicts urbanisation, yellow represents cropping regions and green shows natural environments – note: colours show proportional values of land-use for each pixel with the transparency set as the proportion of each land-use per cell, as such, blended colours represent pixels with mixed land-uses.

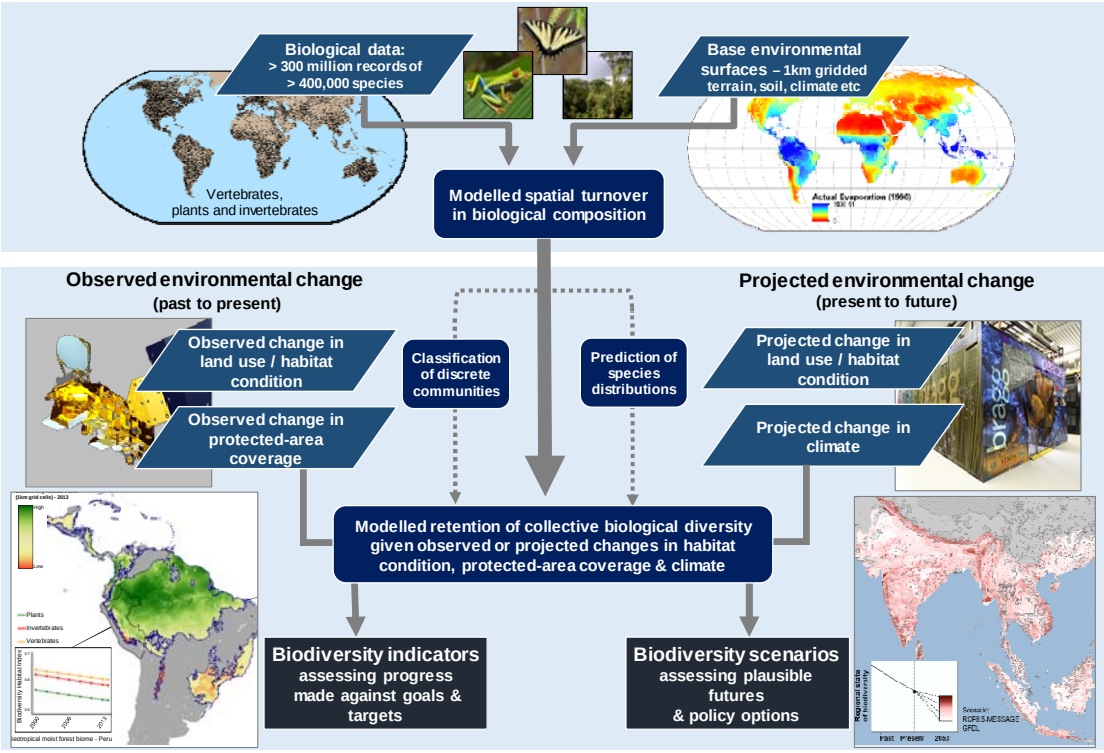

Figure 2: The basic structure of the BILBI modelling framework, showing how our initial modelling of compositional turnover can follow to separate analysis pathways to produce either indicators of recent change in community composition or possible future scenarios of change to composition.

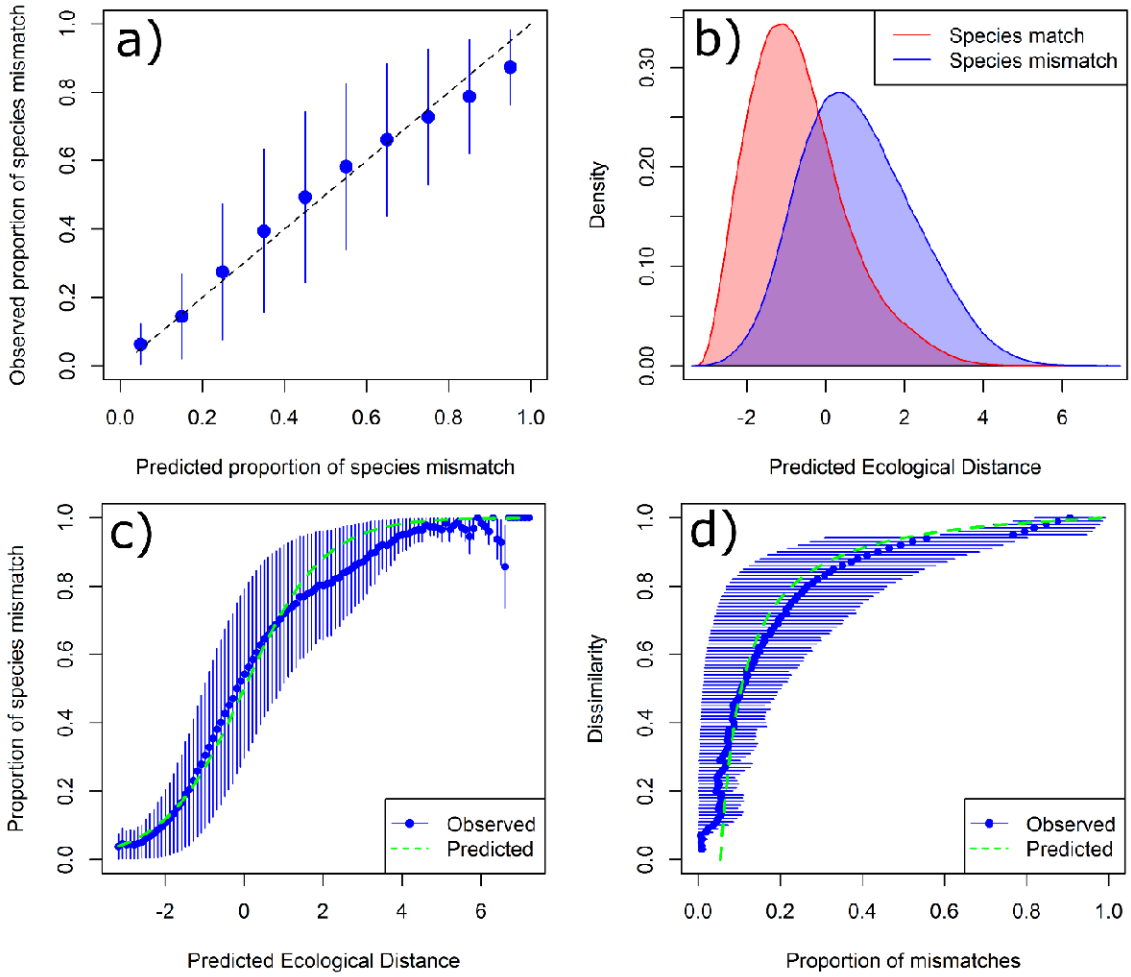

Figure 3: Example model fitting and validation outputs showing: (a) the observed proportion of species mismatches for 10 bins along the predicted proportion of species mismatches from the model. (b) the density of observed species matches and mismatches along the predicted ecological distance (in logit space) from the model. (c) the observed (blue) and predicted (green) proportion of species mismatches for 100 bins along the predicted ecological distance (in logit space). (d) comparison of observed (blue) and predicted (green) compositional dissimilarity against the proportion of mismatches. All error bars show the observed proportion of matches and mismatches ± the variance.

Table 1: Summary counts of the numbers of species and records utilised in the *obs-pair*GDM modelling

	Invertebrates	**Plants**	**Vertebrates**
Number of Species	132761	254145	24442
Number of Records	13244784	52489096	33549534

Table 2: Description of data layers used as covariates during the modelling process

Variable	Short name	Notes	Classification	Source	Reference
Bare ground (proportion)	BARE	G20ESA SoilGrids1km	Soil	www.worldgrids.org	Hengl, T (2005) through ISRIC - WDC Soils
Bulk density	BD	BDRICM SoilGrids1km	Soil	www.soilgrids.org	Hengl, T (2014) through ISRIC - WDC Soils
Clay content mass fraction in %	CLAY	CLYPPT SoilGrids1km	Soil	www.soilgrids.org	Hengl, T (2014) through ISRIC - WDC Soils
pH in H_2O * 10	PH	PHIHOX SoilGrids1km	Soil	www.soilgrids.org	Hengl, T (2014) through ISRIC - WDC Soils
Silt content mass fraction in %	SILT	SLTPPT SoilGrids1km	Soil	www.soilgrids.org	Hengl, T (2014) through ISRIC - WDC Soils
Terrain Ruggedness Index (TRI) GMTED2010	RUG	Mean value of TRI, using Median Statistic, 7.5 arc-seconds. Aggregation done at 1km (0.00833 degree) by calculating the mean values among 16 pixel. Calculated from the GMTED2010 7.5 arc –second product : Data available from the U.S. Geological Survey. (Danielson, 2011)	Terrain	Yale University	(Amatulli *et al.*, 2018)
Topographic Wetness Index	TWI	TWISRE3: SAGA GIS Topographic Wetness Index calculated from SRTM 30+ and ETOPO DEM.	Terrain	www.worldgrids.org	(Hengl, 2013) through ISRIC - WDC Soils
Annual total precipitation (mm)	PTA		Climate	www.worldclim.org	(Hijmans *et al.*, 2005)
Mean minimum temperature of the month of lowest minimum temperature (°C)	TNI		Climate	www.worldclim.org	(Hijmans *et al.*, 2005)
Mean diurnal temperature range of the month of highest diurnal temperature range (°C)	TRX	Radiative adjustment of maximum temperature	Climate (radiative adjustment)	www.worldclim.org	(Hijmans *et al.*, 2005) (Reside *et al.*, 2013b)

Mean maximum temperature of the month of highest maximum temperature (°C)	TXX	Radiative adjustment of maximum temperature	Climate (radiative adjustment)	www.worldclim.org	(Hijmans *et al.*, 2005) (Reside *et al.*, 2013b)
Annual total actual evaporation (mm)	EAAS	Actual evaporation calculated as (Reside *et al.*, 2013b) using Budkyo bucket model. Soil bucket capacity taken as ("Depth to bedrock (R horizon) up to 200cm: www.soilgrids.org" *"Harmonised World Soil Database v1.2: Available Water Content 5km (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012))	Derived climate (radiative adjustment)	www.worldclim.org	(Hijmans *et al.*, 2005) (Reside *et al.*, 2013b)
Mean potential evaporation of the month of minimum potential evaporation (mm)	EPI	Priestley-Taylor evaporation with radiative correction applied to maximum temperature and radiation.	Derived climate (radiative adjustment)	www.worldclim.org	(Hijmans *et al.*, 2005) (Reside *et al.*, 2013b)
Mean water deficit of the month of minimum water deficit (driest) (mm)	WDI	Water deficit calculated as monthly Precipitation-Potential Evaporation so positive water deficit if precipitation exceeds evaporation.	Derived climate (radiative adjustment)	www.worldclim.org	(Hijmans *et al.*, 2005) (Reside *et al.*, 2013b)
Mean water deficit of the month of maximum water deficit (wettest) (mm)	WDX	Water deficit calculated as monthly Precipitation-Potential Evaporation so positive water deficit if precipitation exceeds evaporation.	Derived climate (radiative adjustment)	www.worldclim.org	(Hijmans *et al.*, 2005) (Reside *et al.*, 2013b)

References

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J. & Jetz, W. (2018) A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, **5**, 180040.

Beck, J., Holloway, J.D., Schwanghart, W. & Orme, D. (2013) Undersampling and the measurement of beta diversity. *Methods in Ecology and Evolution*, **4**, 370-382.

Butchart, S.H.M., Clarke, M., Smith, R.J., Sykes, R.E., Scharlemann, J.P.W., Harfoot, M., Buchanan, G.M., Angulo, A., Balmford, A., Bertzky, B., Brooks, T.M., Carpenter, K.E., Comeros-Raynal, M.T., Cornell, J., Ficetola, G.F., Fishpool, L.D.C., Fuller, R.A., Geldmann, J., Harwell, H., Hilton-Taylor, C., Hoffmann, M., Joolia, A., Joppa, L., Kingston, N., May, I., Milam, A., Polidoro, B., Ralph, G., Richman, N., Rondinini, C., Segan, D.B., Skolnik, B., Spalding, M.D., Stuart, S.N., Symes, A., Taylor, J., Visconti, P., Watson, J.E.M., Wood, L. & Burgess, N.D. (2015) Shortfalls and Solutions for Meeting National and Global Conservation Area Targets. *Conservation Letters*, **8**, 329-337.

Calderón-Patrón, J.M., Goyenechea, I., Ortiz-Pulido, R., Castillo-Cerón, J., Manriquez, N., Ramírez-Bautista, A., Rojas-Martínez, A.E., Sánchez-Rojas, G., Zuria, I. & Moreno, C.E. (2016) Beta Diversity in a Highly Heterogeneous Area: Disentangling Species and Taxonomic Dissimilarity for Terrestrial Vertebrates. *PLOS ONE*, **11**, e0160438.

CBD (2010) Strategic plan for biodiversity. In:

Citroen, S., Kempinski, J. & Cullen, Z. (2016) Life after COP21: what does the Paris Agreement mean for forests and biodiversity conservation? *Oryx*, **50**, 201-202.

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

Danielson, J.J.G. & Dean, B. (2011) Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011-1073. In: (ed. U.S.D.O.T. Interior), p. 26. U.S. Geological Survey, Reston, Virginia.

Danielson, J.J.G., Dean B. (2011) "Global Multi-resolution Terrain Elevation Data 2010 (GMTED201O):

U.S. Geological Survey Open-File Report 2011-1073. In: (ed. U.S.D.O.T. Interior), p. 26. U.S. Geological Survey

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677-697.

FAO/IIASA/ISRIC/ISS-CAS/JRC (2012) Harmonised World Soil Database v 1.2. In: (ed. R. Fao, Italy & Iiasa, Laxenburg, Austria)

Ferrier, S. (2002) Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, **51**, 331-363.

Ferrier, S. (2011) Extracting More Value from Biodiversity Change Observations through Integrated Modeling. *Bioscience*, **61**, 96-97.

Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393-404.

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

Ferrier, S., Powell, G.V.N., Richardson, K.S., Manion, G., Overton, J.M., Allnutt, T.F., Cameron, S.E., Mantle, K., Burgess, N.D., Faith, D.P., Lamoreux, J.F., Kier, G., Hijmans, R.J., Funk, V.A., Cassis, G.A., Fisher, B.L., Flemons, P., Lees, D., Lovett, J.C. & Van Rompaey, R.S.A.R. (2004) Mapping more of terrestrial biodiversity for global conservation assessment. *Bioscience*, **54**, 1101-1109.

Francis, A.P. & Currie, D.J. (1998) Globl patterns of tree species richness in moise forests: another look. *Oikos*, **81**, 598-602.

Gray, C.L., Hill, S.L.L., Newbold, T., Hudson, L.N., Börger, L., Contu, S., Hoskins, A.J., Ferrier, S., Purvis, A. & Scharlemann, J.P.W. (2016) Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nature Communications*, **7**, 12306.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. & Townshend, J.R.G. (2013) High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, **342**, 850-853.

Hengl, T.K., M (2013) TWISRE3: SAGA GIS Topographic wetness index. In: (ed. Worldgrids)

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hoskins, A.J., Bush, A., Gilmore, J., Harwood, T., Hudson, L.N., Ware, C., Williams, K.J. & Ferrier, S. (2016) Downscaling land-use data to provide global 30 estimates of five land-use classes. *Ecology and Evolution*, **6**, 3040-3055.

Hudson, L.N., Newbold, T., Contu, S., Hill, S.L.L., Lysenko, I., De Palma, A., Phillips, H.R.P., Senior, R.A., Bennett, D.J., Booth, H., Choimes, A., Correia, D.L.P., Day, J., Echeverría-Londoño, S., Garon, M., Harrison, M.L.K., Ingram, D.J., Jung, M., Kemp, V., Kirkpatrick, L., Martin, C.D., Pan, Y., White, H.J., Aben, J., Abrahamczyk, S., Adum, G.B., Aguilar-Barquero, V., Aizen, M.A., Ancrenaz, M., Arbeláez-Cortés, E., Armbrecht, I., Azhar, B., Azpiroz, A.B., Baeten, L., Báldi, A., Banks, J.E., Barlow, J., Batáry, P., Bates, A.J., Bayne, E.M., Beja, P., Berg, Å., Berry, N.J., Bicknell, J.E., Bihn, J.H., Böhning-Gaese, K., Boekhout, T., Boutin, C., Bouyer, J., Brearley, F.Q., Brito, I., Brunet, J., Buczkowski, G., Buscardo, E., Cabra-García, J., Calviño-Cancela, M., Cameron, S.A., Cancello, E.M., Carrijo, T.F., Carvalho, A.L., Castro, H., Castro-Luna, A.A., Cerda, R., Cerezo, A., Chauvat, M., Clarke, F.M., Cleary, D.F.R., Connop, S.P., D'Aniello, B., da Silva, P.G., Darvill, B., Dauber, J., Dejean, A., Diekötter, T., Dominguez-Haydar, Y., Dormann, C.F., Dumont, B., Dures, S.G., Dynesius, M., Edenius, L., Elek, Z., Entling, M.H., Farwig, N., Fayle, T.M., Felicioli, A., Felton, A.M., Ficetola, G.F., Filgueiras, B.K.C., Fonte, S.J., Fraser, L.H., Fukuda, D., Furlani, D., Ganzhorn, J.U., Garden, J.G., Gheler-Costa, C., Giordani, P., Giordano, S., Gottschalk, M.S., Goulson, D., Gove, A.D., Grogan, J., Hanley, M.E., Hanson, T., Hashim, N.R., Hawes, J.E., Hébert, C., Helden, A.J., Henden, J.-A., Hernández, L., Herzog, F., Higuera-Diaz, D., Hilje, B., Horgan, F.G., Horváth, R., Hylander, K., Isaacs-Cubides, P., Ishitani, M., Jacobs, C.T., Jaramillo, V.J., Jauker, B., Jonsell, M., Jung, T.S., Kapoor, V., Kati, V., Katovai, E., Kessler, M., Knop, E., Kolb, A., Kőrösi, Á., Lachat, T., Lantschner, V., Le Féon, V., LeBuhn, G., Légaré, J.-P., Letcher, S.G., Littlewood, N.A., López-Quintero, C.A., Louhaichi, M., Lövei, G.L., Lucas-Borja, M.E., Luja, V.H., Maeto, K., Magura, T., Mallari, N.A., Marin-Spiotta, E., Marshall, E.J.P., Martínez, E., Mayfield, M.M., Mikusinski, G., Milder, J.C., Miller, J.R., Morales, C.L., Muchane, M.N., Muchane, M., Naidoo, R., Nakamura, A., Naoe, S., Nates-Parra, G., Navarrete Gutierrez, D.A., Neuschulz, E.L., Noreika, N., Norfolk, O., Noriega, J.A., Nöske, N.M., O'Dea, N., Oduro, W., Ofori-Boateng, C., Oke, C.O., Osgathorpe, L.M., Paritsis, J., Parra-H, A., Pelegrin, N., Peres, C.A., Persson, A.S., Petanidou, T., Phalan, B., Philips, T.K., Poveda, K., Power, E.F., Presley, S.J., Proença, V., Quaranta, M., Quintero, C., Redpath-Downing, N.A., Reid, J.L., Reis, Y.T., Ribeiro, D.B., Richardson, B.A., Richardson, M.J., Robles, C.A., Römbke, J., Romero-Duque, L.P., Rosselli, L., Rossiter, S.J., Roulston, T.a.H., Rousseau, L., Sadler, J.P., Sáfián, S., Saldaña-Vázquez, R.A., Samnegård, U., Schüepp, C., Schweiger, O., Sedlock, J.L., Shahabuddin, G., Sheil, D., Silva, F.A.B., Slade, E.M., Smith-Pardo, A.H., Sodhi, N.S., Somarriba, E.J., Sosa, R.A., Stout, J.C., Struebig, M.J., Sung, Y.-H., Threlfall, C.G., Tonietto, R., Tóthmérész, B., Tscharntke, T., Turner, E.C., Tylianakis, J.M., Vanbergen, A.J., Vassilev, K., Verboven, H.A.F., Vergara, C.H., Vergara, P.M., Verhulst, J., Walker, T.R., Wang, Y., Watling, J.I., Wells, K., Williams, C.D., Willig, M.R., Woinarski, J.C.Z., Wolf, J.H.D., Woodcock, B.A., Yu, D.W., Zaitsev, A.S., Collen, B., Ewers, R.M., Mace, G.M., Purves, D.W., Scharlemann, J.P.W. &

Purvis, A. (2014) The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts. *Ecology and Evolution*, **4**, 4701-4735.

Hurlbert, A.H. & Jetz, W. (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences*, **104**, 13384-13389.

Isaac, N.J.B. & Pocock, M.J.O. (2015) Bias and information in biological records. *Biological Journal of the Linnean Society*, **115**, 522-531.

Jenkins, C.N., Pimm, S.L. & Joppa, L.N. (2013) Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, **110**, E2602-E2610.

Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*, **27**, 151-159.

Jones, M.M., Ferrier, S., Condit, R., Manion, G., Aguilar, S. & Perez, R. (2013) Strong congruence in tree and fern community turnover in response to soils and climate in central Panama. *Journal of Ecology*, **101**, 506-516.

König, C., Weigelt, P., Kreft, H. & Baselga, A. (2017) Dissecting global turnover in vascular plants. *Global Ecology and Biogeography*, **26**, 228-242.

Lehner, B. & Doll, P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, **296**, 1-22.

Londoño-Murcia, M.C., Tellez-Valdés, O. & Sánchez-Cordero, V. (2010) Environmental heterogeneity of World Wildlife Fund for Nature ecoregions and implications for conservation in Neotropical biodiversity hotspots. *Environmental Conservation*, **37**, 116-127.

Merow, C., Wilson, A.M., Jetz, W. & Peres-Neto, P. (2017) Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography*, **26**, 243-258.

Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. (2015) Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, **6**, 8221.

Newbold, T., Hudson, L.N., Arnell, A.P., Contu, S., De Palma, A., Ferrier, S., Hill, S.L.L., Hoskins, A.J., Lysenko, I., Phillips, H.R.P., Burton, V.J., Chng, C.W.T., Emerson, S., Gao, D., Pask-Hale, G., Hutton, J., Jung, M., Sanchez-Ortiz, K., Simmons, B.I., Whitmee, S., Zhang, H., Scharlemann, J.P.W. & Purvis, A. (2016) Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science*, **353**, 288-291.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. & Kassem, K.R. (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, **51**, 933-938.

Palma, A.D., Kuhlmann, M., Bugter, R., Ferrier, S., Hoskins, A.J., Potts, S.G., Roberts, S.P.M., Schweiger, O., Purvis, A. & Beggs, J. (2017) Dimensions of biodiversity loss: Spatial mismatch in land-use impacts on species, functional and phylogenetic diversity of European bees. *Diversity and Distributions*, **23**, 1435-1446.

Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R. & Walpole, M. (2010) Scenarios for Global Biodiversity in the 21st Century. *Science*, **330**, 1496-1501.

Ramsay, J.O. (1988) Monotone regression splines in action. *Statistical science*, **3**, 425-441.

Reside, A.E., VanDerWal, J., Phillips, B., Shoo, L.P., Rosauer, D., Anderson, B.J., Welbergen, J., Moritz, C., Ferrier, S., Harwood, T.D., Williams, K.J., Mackey, B., Hugh, S. & Williams, S.E. (2013a) *Climate change refugia for terrestrial biodiversity*. James Cook University.

Reside, A.E., VanDerWal, J., Phillips, B., Shoo, L., Rosauer, D., Anderson, B.A., Welbergen, J., Moritz, C., Ferrier, S., Harwood, T.D., Williams, K.J., Mackey, B., Hugh, S. & Williams, S.E. (2013b)

Climate Change refugia for terrestrial biodiversity: the role of refugia ecosystem resilience and maintenance of terrestrial biodiversity in the face of global climate change. National Climate Change Adaptation Research Facility, Griffith University, Gold Coast, Qld.

Tittensor, D.P., Walpole, M., Hill, S.L.L., Boyce, D.G., Britten, G.L., Burgess, N.D., Butchart, S.H.M., Leadley, P.W., Regan, E.C., Alkemade, R., Baumung, R., Bellard, C., Bouwman, L., Bowles-Newark, N.J., Chenery, A.M., Cheung, W.W.L., Christensen, V., Cooper, H.D., Crowther, A.R., Dixon, M.J.R., Galli, A., Gaveau, V., Gregory, R.D., Gutierrez, N.L., Hirsch, T.L., Höft, R., Januchowski-Hartley, S.R., Karmann, M., Krug, C.B., Leverington, F.J., Loh, J., Lojenga, R.K., Malsch, K., Marques, A., Morgan, D.H.W., Mumby, P.J., Newbold, T., Noonan-Mooney, K., Pagad, S.N., Parks, B.C., Pereira, H.M., Robertson, T., Rondinini, C., Santini, L., Scharlemann, J.P.W., Schindler, S., Sumaila, U.R., Teh, L.S.L., van Kolck, J., Visconti, P. & Ye, Y. (2014) A mid-term analysis of progress toward international biodiversity targets. *Science*, **346**, 241-244.

UN (2015) Transforming our world: the 2030 agenda for sustainable development. In: (ed. U. Nations), New York.

UNEP & WCMC (2016) Protected planet: Teh World Database on Protected Areas. In. UNEP-WCMC, Cambridge, UK.

Watson, J.E.M., Jones, K.R., Fuller, R.A., Marco, M.D., Segan, D.B., Butchart, S.H.M., Allan, J.R., McDonald-Madden, E. & Venter, O. (2016) Persistent Disparities between Recent Rates of Habitat Conversion and Protection and Implications for Future Global Conservation Targets. *Conservation Letters*, **9**, 413-421.