Representation of global change drivers across biodiversity datasets

Authors: Gergana N. Daskalova[1]*, Diana E. Bowler[2-4], Isla H. Myers-Smith[1], Maria Dornelas[5]

Affiliations:

[1] School of GeoSciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3FF, Scotland

[2] German Centre for Integrative Biodiversity Research (iDiv), Puschstrasse 4, Leipzig 04103, Germany

[3] Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger Straße 159, 07743 Jena, Germany

[4] Helmholtz Center for Environmental Research - UFZ, Department of Ecosystem Services, Permoserstraße 15, 04318 Leipzig, Germany

[5] Centre for Biological Diversity, University of St Andrews, Greenside Place, St Andrews KY16 9TF, Scotland

***Correspondence to:**

Gergana Daskalova, Room 401, Crew Building, School of GeoSciences
University of Edinburgh, West Mains Road, Scotland
gndaskalova@gmail.com

Abstract

Global change has altered biodiversity and impacted ecosystem functions and services around the planet. Understanding the effects of anthropogenic drivers like human use and climate change on biodiversity change has become a key challenge for science and policy. However, our knowledge of biodiversity change is limited by the available data and their biases. Over land and sea, we test the representation of three worldwide and multi-taxa biodiversity databases (Living Planet, BioTIME and PREDICTS) across spatial and temporal variation in global change and across the tree of life. We find that variation in global change drivers is better captured over space than over time around the world and across the previous 150 years. Spatial representation of global change was as high as 78% in the marine realm and 31% on land. Our findings suggest ways to improve the use of existing biodiversity data and better target future ecological monitoring.

One sentence summary

Biodiversity data capture most of the variation in global change but filling the remaining data gaps will allow us to better understand ongoing change and predict future trajectories for Earth's biota.

Introduction

Human activities are reshaping the planet from the tropics to the poles and across land and sea[1–3], and the Earth's biodiversity is shifting in response[4]. Parallel with this rapid biotic reorganization, an ecological data revolution is underway with more open-access data available now than ever before[5,6]. Large-scale data compilations (e.g., Living Planet[7], BioTIME[8], PREDICTS[9], GBIF[10], TetraDensity[11]) have been analyzed to test general patterns of biodiversity change across the world and under impacts of anthropogenic drivers[4,12–16]. Such studies have revealed a wide spectrum of biodiversity change, including both increases and decreases of species richness and abundance, with trends quantified over time[17–20], space[16,21] and taxa[22,23]. The biodiversity data underlying many of these syntheses (e.g., time series, occurrence records and space-for-time surveys) have already been shown to be biased geographically and taxonomically[24–27]. Surprisingly, much less attention has been given to whether the data are also biased with respect to the overall variation in global change drivers. Yet, knowing to what degree the sampling of biodiversity databases captures global change is vital for interpreting results derived from data syntheses and identifying future data gaps to be filled. The next stage of biodiversity syntheses, scenarios and conservation goals will be brought together in the Convention on Biological Diversity's Post-2020 Global Biodiversity Framework. We argue that post-2020, biodiversity science needs to move towards improved representation in biodiversity data, including the heterogeneous distribution and sampling of global change.

Insights from large-scale data syntheses inform trajectories of past, current and future change in the Earth's biota[4,7,28,29], as well as the development of indicators for global conservation policies[30,31]. To upscale the findings of syntheses of local-scale data to estimates of global or mean biodiversity change, the underlying data should be representative across multiple dimensions: space, time, taxonomic variation, as well as variation in the drivers of biodiversity change[32–35]. There is already recognition of the biases associated with space and geography[24,26,32,34], time and historical baselines[24,26,36,37] and taxonomy[33,37,38]. In contrast, sampling biases with respect to drivers of change are rarely emphasized in the existing literature (but see [39] for spatio-taxonomic biases in North American butterfly occurrence records). At smaller spatial scales, sampling biases are well-documented in national monitoring schemes and citizen science data (e.g., showing over-representation of urban areas[40] or under-representation of regions undergoing rapid climate change[39]). In contrast, at the larger spatial scales of data syntheses, sampling biases associated with global change drivers remain unknown. Such knowledge gaps compromise our ability to draw broad inferences from the outcomes of syntheses and to quantify the shape of the relationship between driver intensity, such as extent of land-use change, and biodiversity, in order to identify ecological tipping points[35,41–43]. Thus, to interpret the findings of any data synthesis, we need to consider if the underlying data are sampled from sites with the full range of different driver intensities, or rather mostly include heavily impacted sites or intact wilderness areas[4,35,44]. Understanding the representativeness of biodiversity data across global change axes is essential to interpret estimates of regional or global-scale biodiversity change from compilations of local-scale data.

Here, we quantify global change representation in biodiversity databases, present our perspective on capturing the representativeness of biodiversity data in large-scale syntheses and discuss implications for interpreting their findings. We focus on four aspects of representativeness – global change intensity over space, global change intensity over time, geography, and taxonomy. Particularly, we highlight the patterns in spatial and temporal sampling of global change drivers as under-explored types of bias. Our perspective serves two important purposes: 1) to highlight the variation in global change drivers that is already captured by global datasets and hence the driver impacts that we can quantify in ongoing studies, and 2) to identify the gaps in data representativeness that future studies, monitoring and data mobilization actions should target. By building biodiversity databases that are more representative of multiple axes of natural and anthropogenic variation, we can improve predictions of the global state and trends of biodiversity.

We combined three of the largest, currently existing, open-access biodiversity databases (Living Planet[7] - marine and terrestrial; BioTIME[8] - marine and terrestrial; and PREDICTS[9] - terrestrial) with maps of global change drivers[3]. To measure how well each database captures variation in global change intensity over space, we first estimated driver variation around the world. We used data indicating the five big drivers of climate change, human use, human population density, pollution and invasive species pressure[3] - and determined the ‘global change space’ using the dominant orthogonal axes of change (similar to the concept of trait or niche space[45]). We then mapped the sampling sites within each database onto the global change space to highlight the sampled region, as well as regions with under- or over- representation. To quantify the representation of global change intensity over time, we focused on climate change and land cover change across terrestrial sites with time series data from the Living Planet and BioTIME databases. At each site, we compared the amount of change that occurred before relative to during the periods of biodiversity monitoring. To estimate geographic representation, we mapped site locations of the three databases and determined sampling intensity across a spatial grid covering the planet. Finally, to estimate taxonomic representation, we calculated the percentage of known species included in the three databases.

Biodiversity data capture spatial variation in global change space at sea, but not on land

Overall, we found that biodiversity data from the Living Planet, BioTIME and PREDICTS databases capture a surprisingly high amount of the spatial variation in global change intensity around the planet, especially in the marine realm (Figure 1). There was between 71% and 78% overlap between global change space and the variation sampled by biodiversity databases in the marine realm, versus 17% to 31% in the terrestrial realm (Figure 1). Among the five global change drivers we tested, climate change and pollution in the marine realm were sampled the most representatively, suggesting that we can test the effects of these drivers with higher confidence and the underlying data could be used when creating global scenarios for the future (Figure 2, Extended Data Table S1). The terrestrial global change space was less well sampled and the highest overlap with global change was 31% for the Living Planet Database (Figure 1). In fact, all three databases predominantly sampled places with medium to high human use and lacked data from regions with low land-use change and pollution. Similarly, across both realms, but particularly strongly over land, all databases were lacking sites that have experienced high amounts of climate change, reflecting geographic gaps in data collection in places like the Arctic (Figure 4). Following experimental design principles, manipulative studies to determine treatment effects often include a range of treatment levels from low to high in order to have sufficient statistical power[46]. We propose extending experimental design thinking to syntheses of observation studies that aim to attribute change to a driver by ensuring data are included from sites experiencing a range of driver intensities.

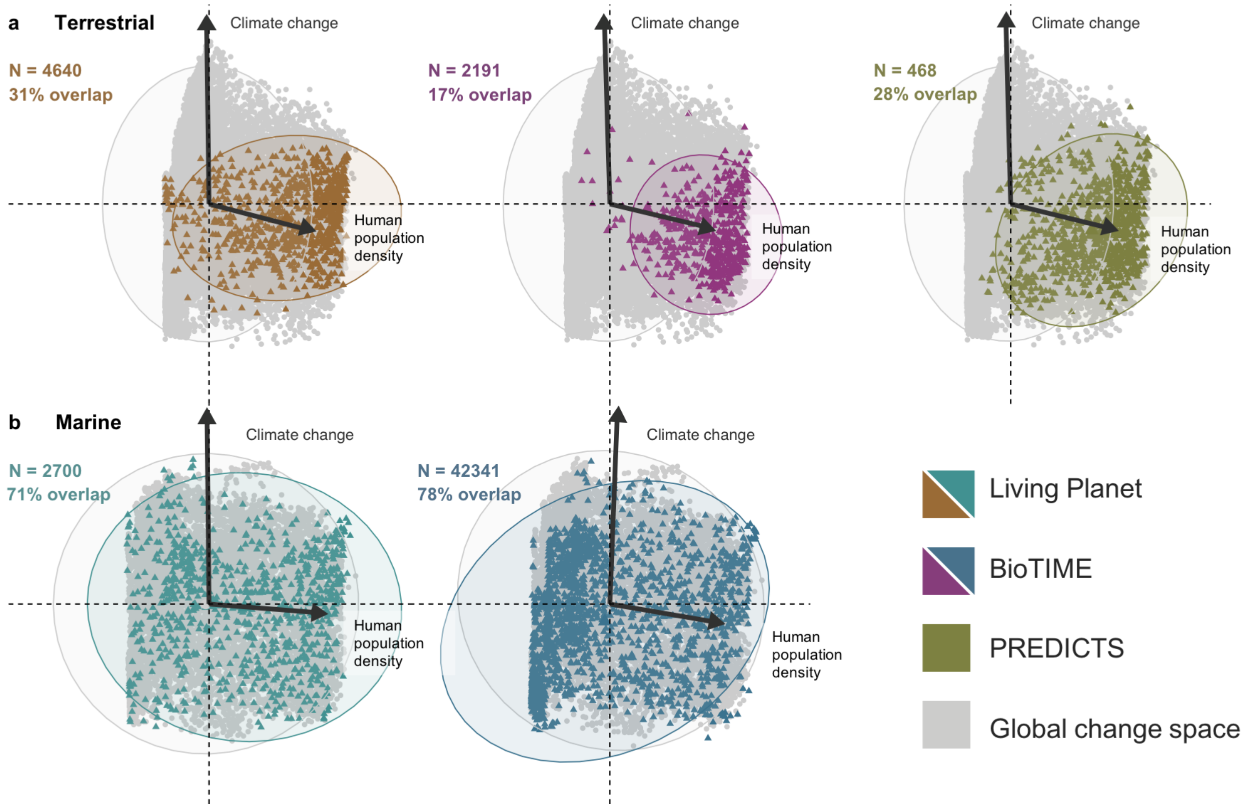

Figure 1. Biodiversity data capture spatial variation in global change space better in the marine versus terrestrial realm. Figure shows Principal Component Analysis of the terrestrial

(panel **a**) and marine (panel **b**) magnitudes of human use, climate change, human population density, pollution and invasion potential across the locations of the Living Planet, BioTIME and PREDICTS databases as well as one million randomly sampled locations across the full extent of the globe (in grey). PCA axes omitted for visual clarity. Arrows show direction and magnitude of PCA scores. Human use, pollution and invasion potential were correlated with human population density. For details on the global change driver layers, see Bowler et al. 2020. Annotations show sample size (N) and the percentage overlap between the 95% prediction ellipses covered by random sampling of global change space and the variation in global change sampled by the different databases.

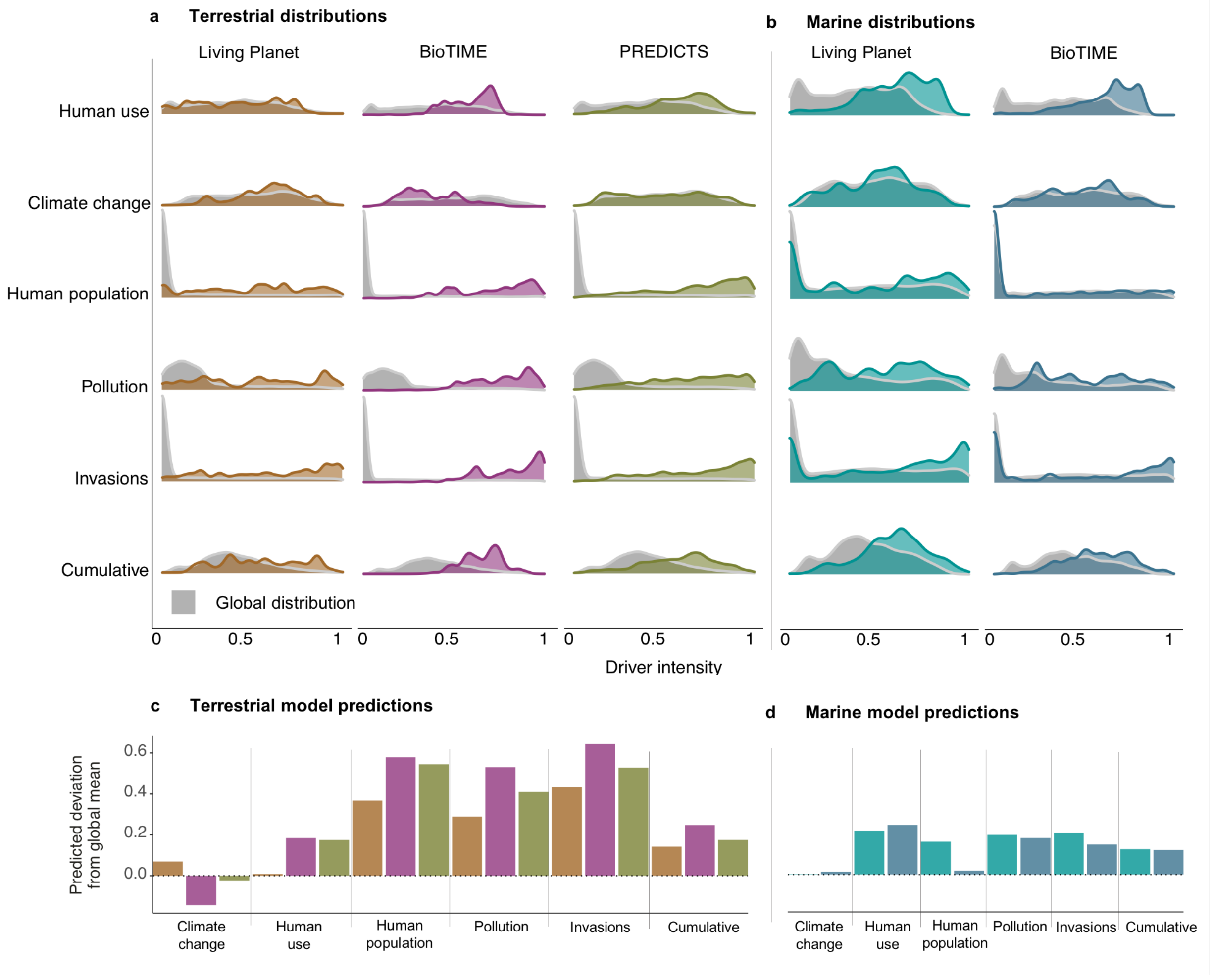

Figure 2. Higher magnitudes of global change drivers are overrepresented in biodiversity data. Panels **a** and **b** show distributions of the raw global change driver data from random sampling spanning the globe and sites from existing biodiversity databases. Panels **c** and **d** show effect sizes of general linear models comparing the magnitude of global change drivers (response variable) across the Living Planet, BioTIME and PREDICTS databases and a random sampling of the planet (categorical explanatory variable). Positive effect sizes indicate higher average magnitudes at the sampled sites within databases than in the random global sampling, and negative effect sizes indicate lower average magnitudes. Because of the large sample sizes included in the statistical models, the 95% credible intervals around the effect sizes were too small to be visualized in the figure.

Biodiversity data often miss the temporal peaks of land cover change, but capture those of climate change

We found mismatches between when global change occurred and the timing of biodiversity data collection, which were more frequent for land-use change than for climate change (Figure 3). While it is well-known that peak land-use conversion often predates ecological monitoring by centuries to millennia (e.g., [1,47,48]), studies rarely quantify the magnitude of this mismatch or account for the long-term trajectory or historic baseline (but see [13,21]). For drivers such as forest loss, the peak often occurred decades to centuries before the start of most biodiversity monitoring (Figure 3a-b, [13]). In contrast, for climate warming, a driver that is more pronounced in more recent decades, we found that the majority of the Living Planet and BioTIME time series (76% and 56% of terrestrial time series, and 64% and 59% of marine time series, respectively) have experienced larger magnitudes of warming during the period of monitoring when compared to the same length of time preceding data collection. Thus, biodiversity data better captured contemporary warming relative to other global change drivers (Figure 3c-f). Our results suggest that weaker or stronger relationships between biodiversity time series and drivers such as forest loss and climate change likely reflect differences in the time periods when each driver was most intense.

The sampled variation in global change driver intensity over time can influence the strength of relationships detected in attribution analyses[13,35,43] and can obscure assessment of biodiversity trends in ecosystems with tipping points[49]. Monitoring schemes that start well after the peak magnitude of a global change driver will likely underestimate that driver's impact on biodiversity[35]. Equally, lagged biodiversity change might mean that the effects of land-use drivers like forestry or agriculture persist decades after harvest or farming has ceased[13,43]. These interactions between lagged biodiversity responses to disturbance and temporal variability of global change have produced heterogeneous and often non-linear biodiversity trends, as have been reported for many taxa, including birds[50], moths[20] and wasps[51]. Additionally, analyses of observational datasets with both short durations and little variation in global change intensity over time have reduced statistical power and thus might fail to detect the effect of global change drivers[52]. The temporal mismatch of ecological monitoring and global change drivers is hard to alleviate because new data collection cannot fill historic data gaps. To move forward, we suggest mobilizing as much existing data as possible and improving data accessibility, developing methods to infer data we cannot observe and including baselines and variation in driver intensity over time in statistical models.

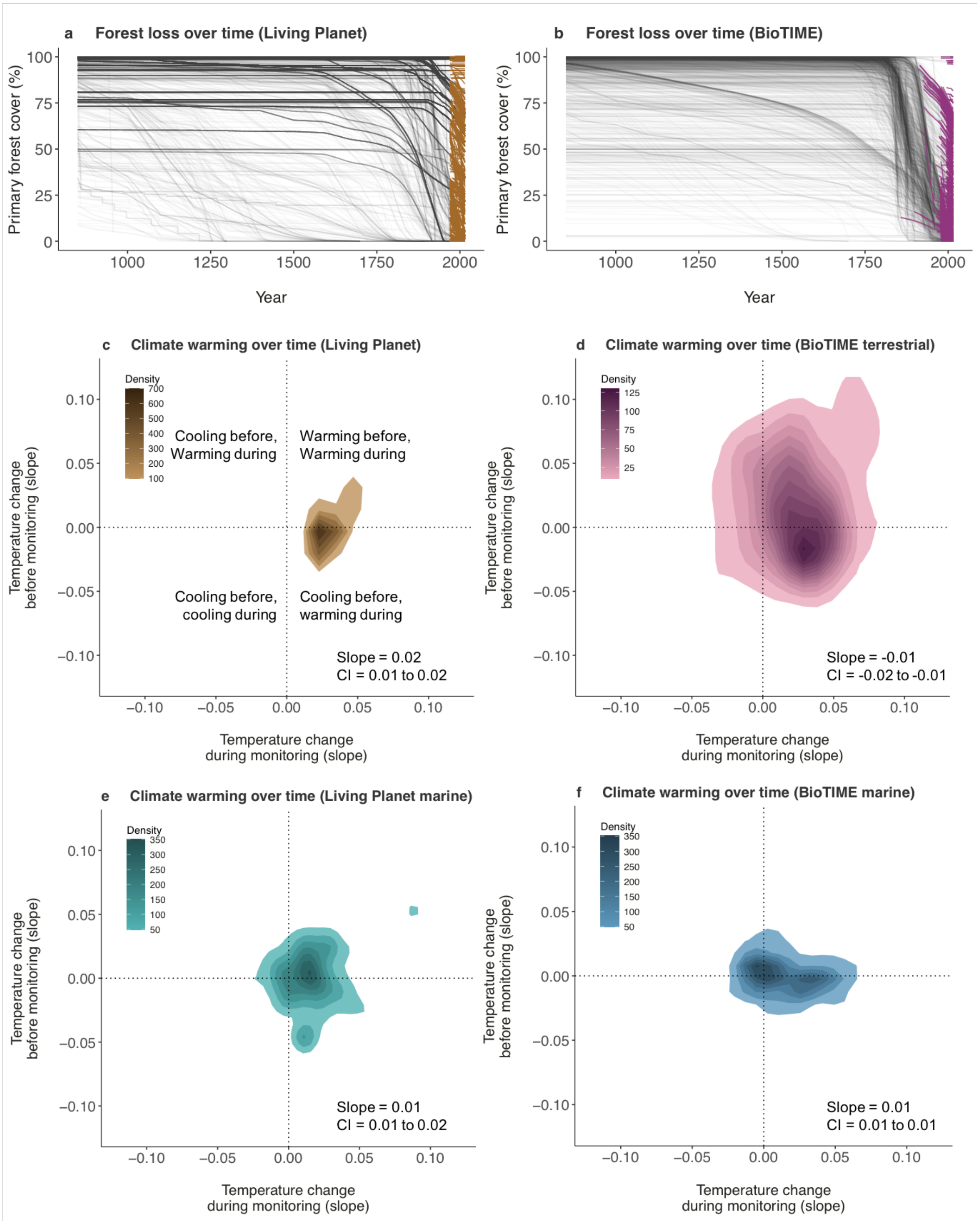

Figure 3. The majority of primary forest was lost by the time ecological monitoring began whereas high magnitudes of climate warming predominantly occurred during the time

series. Panels **a** and **b** show the temporal trajectory of primary forest loss across sites part of the Living Planet (N = 4640) and BioTIME (N = 2191) databases. The primary forest cover estimates show proportions based on the LUH database[47] and were calculated for cells of approximately ~96km^2 around the centerpoint of each site. Historic human use time series data of sufficient duration were not available for the marine realm. The periods for comparison in panels **c**-**e** were the same as the duration of each time series and were always more than five years (for example for a time series starting in 2000 and ending in 2010, we used 1990-2000 as the comparison period). Slope values on the axes of panels **c**-**e** show changes in temperature in degrees Celsius per year, derived from general linear models estimating temperature as a function of year. For the terrestrial realm, surface air temperature was obtained from the CRU TS v4.05 database[53] and for the marine realm, the sea surface temperature data was extracted from the NOAA Extended Reconstructed SST v5 database[54]. Slope and credible interval annotations on panels **c**-**e** show the posterior mean for the average temperature change in the period during monitoring relative to before monitoring.

Geographic gaps in biodiversity data do not always result in gaps in global change space

Underrepresentation in geographic space did not directly translate into gaps in global change space and thus an incomplete geographic sample can capture a surprising amount of variation in global change driver intensity (Figures 1-2, 4). Geographic gaps exist across all three databases we tested, particularly in tropical and high latitudes and in the deep sea. Regions including Northern Asia, Africa and South America had fewer sample sites than Europe and North America across all three databases. For example, there were twice as many European records as there were South American ones in the PREDICTS database, despite South America being almost twice the size of Europe. Europe and North America not only had more sampling across space, but repeat sampling was also more frequent (Figure 4a-e). Ecoregions in the marine realm were better represented than those in the terrestrial realm, with data sampled in 69% and 48% of marine ecoregions in the Living Planet and BioTIME databases, compared with the same in 16%, 30% and 32% of terrestrial ecoregions in the Living Planet, BioTIME and PREDICTS databases, respectively (Figure 4). Geographic biases are well-known caveats of biodiversity data (e.g., [24,26,55]) and can be particularly problematic when extrapolating from patchy local-scale data to broad macroecological patterns[4]. For example, studies of insect biodiversity trends from a limited sample of geographic locations have found steep declines (e.g. 63 sites in Germany and 73 sites in predominantly North America and Europe in [56,57], respectively), whereas studies from larger and more geographically representative datasets have found no net change[58]. This nuance around the source locations of biodiversity data is often lost in media and public communication of population and biodiversity change, sometimes leading to misinterpretation of local declines as ubiquitous worldwide[59–62]. We suggest targeting future ecological monitoring to fill in the gaps in not just geographic but also global change space to better capture and communicate the variety of ways in which humans are altering biodiversity around the world.

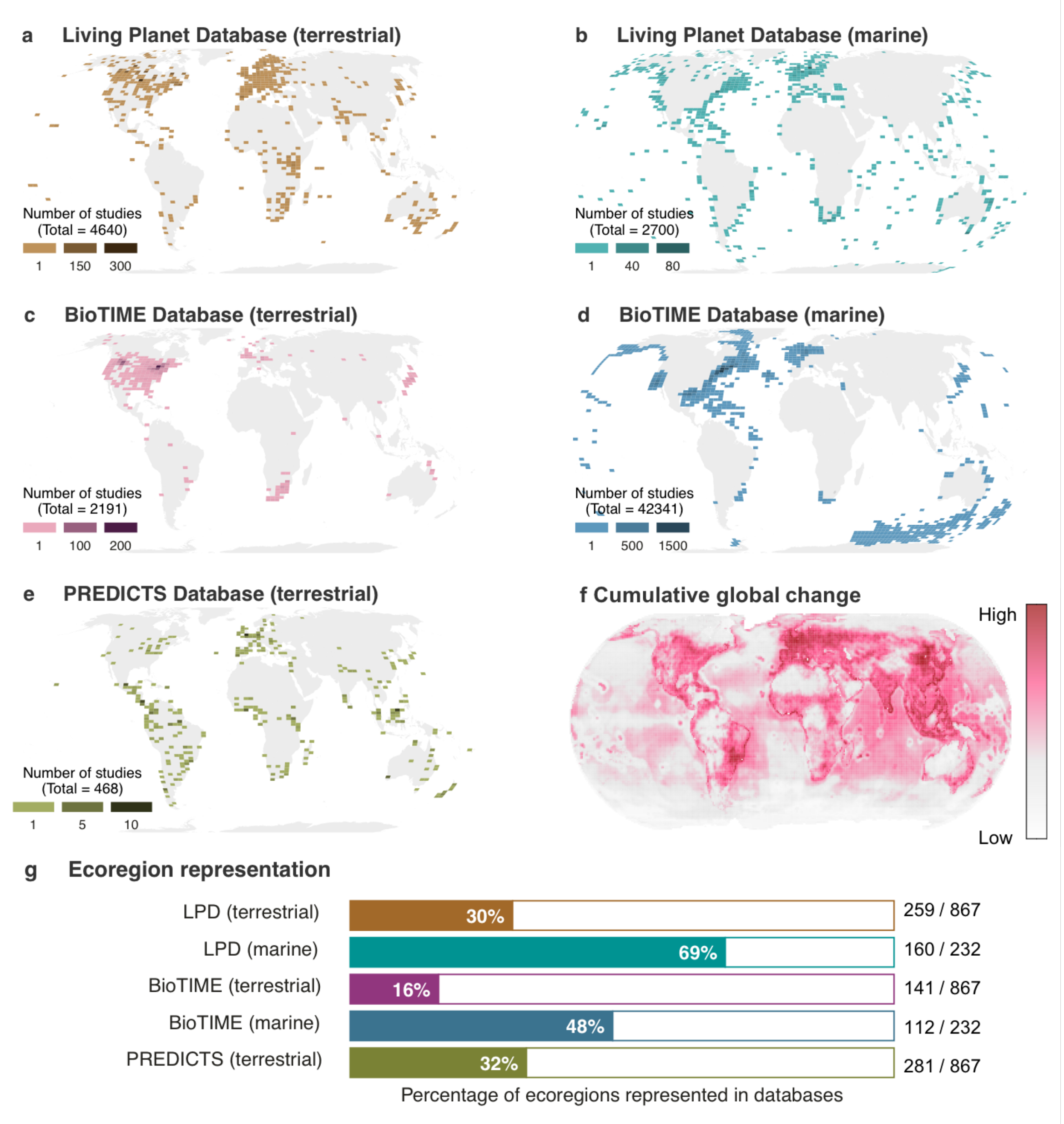

Figure 4. Geographic and ecoregion gaps in biodiversity data exist in both the marine and terrestrial realms but they do not directly translate to gaps in global change variation. Maps on panels **a**-**e** show locations of sites from the Living Planet, BioTIME and PREDICTS databases with darker colors indicating higher numbers of sites. Panel **f** shows the intensity of cumulative global change (climate change, human use, human population density, pollution and invasion pressure combined) across the terrestrial and marine realms, based on[3]. Ecoregions are based

on the classification of [63]. Number annotations on panel **f** show the number of ecoregions represented by at least one record and the total number of marine and terrestrial ecoregions on Earth.

More and less well represented taxa may respond differently to global change drivers

Taxonomic representation in biodiversity analyses could influence the detected global change responses, with certain taxa being more or less sensitive to global change[64–67]. For example, longer-lived species may have greater lagged responses to global change drivers such as land-use change when compared with species with shorter generation times[13]. We found that birds were the best-, and arthropods the worst-represented taxa across the Living Planet, BioTIME and PREDICTS databases (Figure 5), as commonly found in ecological datasets[8,9,65]. Recently, invertebrates and in particular insects have been highlighted as a taxon experiencing potential steep declines in abundance and biomass[56,57], yet such findings are confounded by the general paucity of invertebrate data[60–62] (but see [68] for a recent effort in compiling insect data). In contrast, birds are the focus of many national and international monitoring schemes and for many species, research has established how populations are changing over time[69]. Concurrently, there are frequent calls for better sampling across the tree of life to capture the variety of ways in which species from the smallest ant to the biggest sequoias are being impacted by the Anthropocene (e.g.,[70–76]). Without representative taxonomic coverage, we could be failing to characterize the full balance between the winners and losers of particular global change driver[77]. While our findings show that the spatial variation in global change is broadly well-sampled by the three databases we tested (Figure 1), it is important to highlight that the majority of those biodiversity records are for mammals, birds and plants. Consequently, global change space remains poorly represented for less studied taxa like terrestrial invertebrates for which representation was only 3.2% for time-series data (BioTIME) and 29.4% for space-for-time data (PREDICTS, Extended

Data Figure 1). Extending findings from the limited representation of the planet's diversity to cross-taxa scenarios of future change should be done with caution and placed in the context of which species have the most records within the database[78–80].

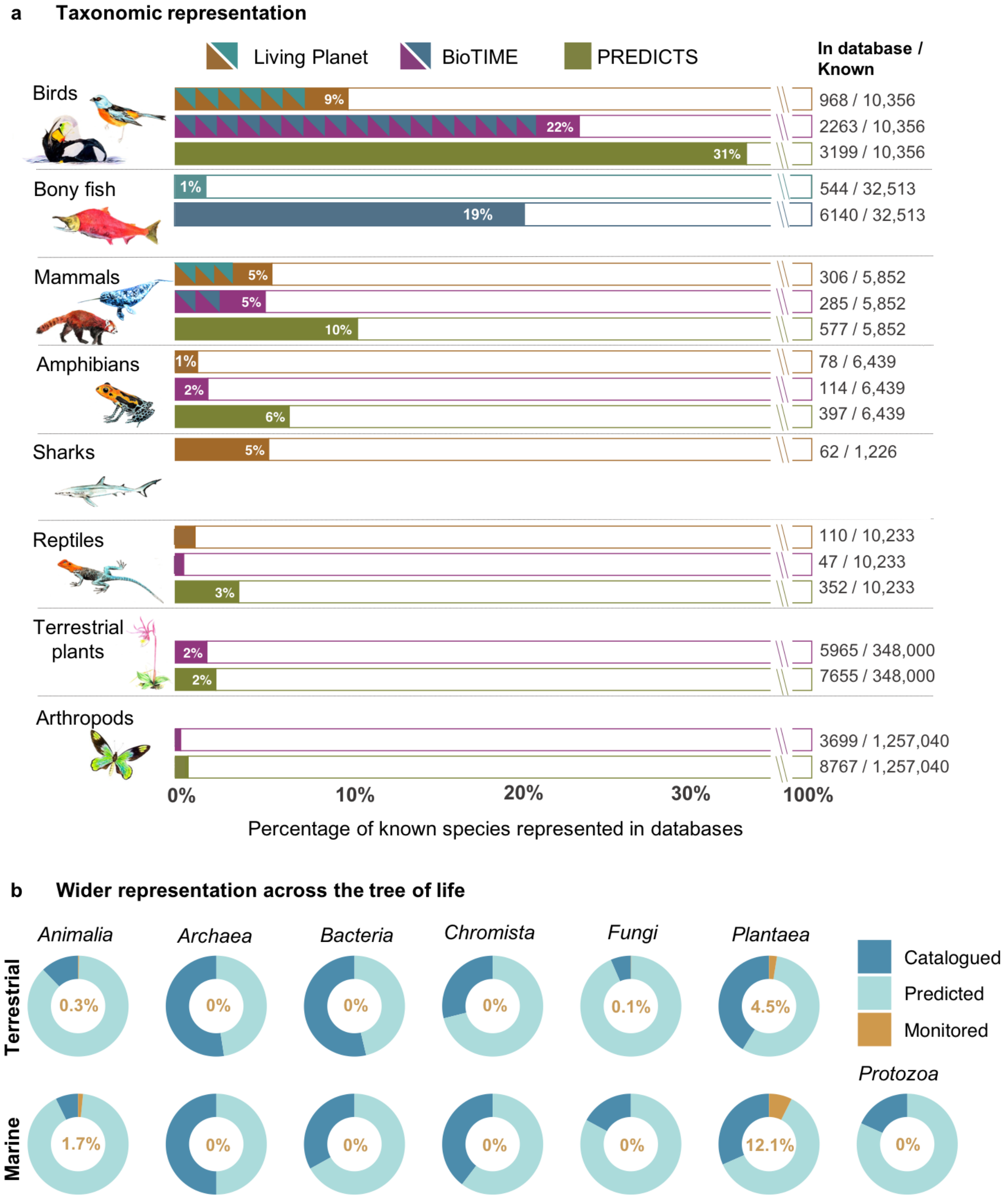

Figure 5. Taxonomic representation of biodiversity data is highest for birds and mammals and lowest for arthropods. The data available across biodiversity databases do not reflect the taxonomic diversity of the tree of life and millions of species are not represented by even a single record (**b**). Percentages in **a** show how many of the known species in each taxon are represented by at least one record in the Living Planet, BioTIME and PREDICTS databases. Panel **b** shows how monitored species fit within the larger tree of life and is based on catalogued and predicted species in [81]. The "Monitored" category combines the species represented in the Living Planet, BioTIME and PREDICTS databases and the percentages show how many of the predicted species feature at least once in biodiversity databases. The numbers of known species per taxa were extracted from the 2019 edition of the Catalogue of Life (http://www.catalogueoflife.org). The values for the birds and mammals in the Living Planet and BioTIME database include both marine and terrestrial species. Note that the BioTIME database additionally include records for marine invertebrates, benthos, marine plants, freshwater plants, freshwater invertebrates and freshwater fish.

Recommendations for capturing the spectrum and distribution of global change across space, time and the tree of life

Understanding ongoing and future biodiversity change can be improved by quantitatively accounting for the representation of biodiversity data across global change space, over the temporal trajectory of drivers, across geographic regions and across the tree of life. Together, our four recommendations provide guidance on using existing observational data, determining where to locate future ecological monitoring and designing experimental studies of novel global change space without modern day analogues.

Recommendation 1: Test the global change representation of databases and syntheses

Extending our thinking beyond just geographic, temporal and taxonomic bias to include global change variation can contextualize research findings from biodiversity data. The different relative positions of the current forms of global biodiversity databases within global change space might explain some of the differences in research findings. For example, predominantly negative impacts of intensifying land-use change have been found using PREDICTS[16], both negative and positive influences of forest loss based on Living Planet and BioTIME[13] and stronger impacts of temperature change on richness, composition and abundance trends in BioTIME[14]. In this study, we present a framework to test data representation across different global change drivers over space and time that can be applied to other datasets. We recommend that future syntheses explicitly include tests of the representation of their data for the global change drivers being tested in addition to highlighting other data gap[26,33,38,65,71–74].

Recommendation 2: Account for data representation across multiple axes in existing syntheses of observational data

Beyond testing for global change representation, studies should ideally account for the representation of their data for the global change driver(s) of interest. A variety of approaches could be used, including the following: 1) Randomized subsampling can help balance uneven data where certain types of global change are overrepresented while others are underrepresented[37], however, this has the disadvantage of discarding potentially valuable data. 2) Statistical weightings have been used to adjust the representativeness of the data sample e.g., by up-weighting under-represented regions or taxa (e.g., as employed by the Living Planet Index[82] and often with citizen science data[31,73]) but this approach can over emphasize the effect of very small portions of the overall data[83] and potentially inflate errors associated with those data[36,60,83,84]. 3) Bias can be explicitly modelled using fixed effects for continuous variables of driver intensity and random effects to represent geographic, temporal and taxonomic structure (e.g., as in [85]), but care must be taken to ensure all uncertainties are propagated through to the global mean estimate[86–89]. 4) Baselines, time since disturbance and changing intensity of impact of global change drivers can be explicitly incorporated into analyses of time series data[13,43]. Analyses that explicitly incorporate global change representation will provide more accurate attribution of biodiversity change to global change drivers.

Recommendation 3: Prioritize new data collection for underrepresented parts of the global change spectrum

A lot of the focus in the literature is on filling geographic[24,26,32,34], temporal[24,26,36,37], and taxonomic[33,37,38] biodiversity data gaps, but this focus should be shifted towards prioritizing regions that undersample global change. For example, we are currently lacking biodiversity data from places with high magnitudes of climate change including Arctic and boreal forest regions, as well as tropical regions that are currently entering non-analog climate spac[90]. These data are important not just for understanding current effects of climate change, but also as sentinels of future change around the world[90–92]. Another underrepresented part of the global change spectrum is relatively intact sites with low human impact (Figures 1-2), which provide a necessary comparator for testing the impacts of human use, pollution and other global change drivers. Although we cannot achieve greater global change representation of historic and current data, the monitoring programs of the future can prioritize global change representation, while also filling geographic and taxonomic gaps.

Recommendation 4: Design experiments to study novel global change space

Global change space is not static and to make scenarios for future biodiversity trends, we need to sample not only current variation in global change drivers, but also future combinations of global change drivers[93]. We suggest that using projections for climate change and human impact, such as IPCC[2] and HYDE[48], we can compute future global change space and determine novel environments without current-day analogues and where those novel environments will most likely occur. Designing lab and field experiments that test novel combinations and magnitudes of global change drivers can provide a preview of biodiversity responses to future environmental conditions. Prioritizing biodiversity monitoring where novel environments will likely develop will ensure that future biodiversity syntheses and impact assessments will represent future as well as current global change.

Conclusion and ways forward

Predicting future biodiversity change and its consequences for ecosystem functions and services to society is an urgent scientific challenge. Global biodiversity monitoring needs to capture a representative sample of the world over both space and time, as well as the full spectrum of global change drivers. In this study, we quantified four types of representativeness - global change intensity over space, global change intensity over time, geography, and taxonomy (Figures 1-5). Together, our findings demonstrate that global biodiversity datasets capture a large proportion of the intensity of global change, but not uniformly. Over space, existing data capture up to 78% of the spatial variance in global change drivers, but more so at sea than on land (78% versus 31%). Over time, monitoring often starts after the peak intensity in environmental change for drivers like primary forest loss[13], but more closely coincides with the period of rapid climate change (Figure 3). We identify four recommendations to test and account for current and future global change representation: 1) test the global change representation of databases and syntheses, 2) Account for data representation across multiple axes in existing syntheses of observational data, 3) Prioritize new data collection for underrepresented parts of the global change spectrum, and 4) Design experiments to study novel global change space.

The biodiversity synthesis literature must progress beyond merely discussing bias to instead quantify and account for the global change representation of biodiversity data. By considering all axes of the global change spectrum, we can strengthen the empirical evidence for the next stage of IPBES global biodiversity assessments and the global biodiversity indicators for the Post-2020 Global Biodiversity Framework. With continued calls for more biodiversity data (e.g.,[4,94,95]), we especially advocate for future biodiversity monitoring to target not just geographic and taxonomic gaps, but also improved representation of global change.

Methods

Databases of ecological monitoring

We combined three of the largest biodiversity databases - Living Planet (7,340 time series spanning 1970-2014), BioTIME (44,532 time series spanning 1858-2017) and PREDICTS (468 studies spanning 1984-2013). The Living Planet database[7] includes time series data of individual species' abundance for vertebrate taxa for the terrestrial, marine and freshwater realms (freshwater realm data were excluded for the purposes of this analysis because of lack of global change driver data for freshwater environments). The BioTIME database[8] is also a compilation of time series but of ecological assemblages for vertebrate, invertebrate and plant taxa across the marine and terrestrial realms. The PREDICTS database[9] includes space-for-time comparison studies testing the effects of land-use change on vertebrates, invertebrates and plants and thus focuses on the terrestrial realm.

Databases of global change

We used the 16 marine and terrestrial global change driver layers compiled by Bowler et al. 2020[3] (Extended Data Table S2). We selected these layers because they had been harmonized across both realms and hence were most suitable for our global analysis. As in Bowler et al., these layers were grouped into five focal drivers: human use (land-use for the terrestrial realm, and exploitation for the marine realm), climate change, human population density, pollution and invasion potential. The driver data were harmonized to a standard spatial grid with a resolution of 100 km^2 and were aggregated over the time period between 1990 and 2010. With the exception of forest loss and climate change, the driver data were not available on an annual time step. Data limitations are particularly pronounced for the marine realm, as it is harder to monitor global change at sea than over land. For details on the individual layers forming the global change data, including their resolutions and temporal coverage, see Extended Data Table S2. We used the Land Use Harmonisation (LUH) database of reconstructed historical land cover at a 0.25° resolution[47] to extract primary forest cover estimates over a long historic period (from the year 800 to 2014). For the terrestrial realm, we obtained monthly surface air temperature at a spatial resolution of 0.5° from the CRU TS v4.05 database[53] and for the marine realm, we extracted sea surface temperature at a spatial resolution of 2° from the NOAA Extended Reconstructed SST v5 database[54]. For both surface air temperature and sea surface temperature, we aggregated the monthly data into yearly averages for time periods matching the timing of biodiversity time series as well as the period of same duration preceding the monitoring (e.g., for a time series from 2000 to 2010, we extracted data from 1990 to 2000 and from 2000 to 2010).

Mapping ecological monitoring in global change space

We combined the geographical coordinates of all spatially-explicit monitoring sites in the Living Planet, BioTIME and PREDICTS databases. For each sampling site, we extracted the intensity of 16 global change layers as well as their cumulative magnitudes. The driver data matching the sites in each database are available in an open-access repository (see Code and Data Availability section). The estimates for the magnitudes of each driver were standardized between 0 and 1 to make them comparable. We used a Principal Component Analysis (PCA) to map global change space within the two dominant orthogonal axes (similar to trait space[45]), which explained 81% of the variation, and visualized the sampled sites in this global change space. We extracted driver intensity for one million simulated random locations spanning the globe to represent an unbiased sample of the marine and terrestrial surface of the world. We used this random sample as a comparison for quantifying the representation of global change variation in biodiversity data. To calculate the percentage overlap between global change space and the area within it occupied by the three databases, we used the package SIBER v.2.1.6.9[96] and 95% prediction ellipses. The overlap was calculated using ellipses based on the climate change and human use variables, since human population density, pollution and invasion pressure were positively correlated with human use. We visualized marine and terrestrial global change space separately because of known differences in the global change driver variables capturing human impact across realms and suspected differences in the patterns of sampling effort[3].

To statistically compare the intensity of global change drivers around the world and in locations with biodiversity data, we used two Bayesian general linear models (one for the marine and one for the terrestrial realm) with driver intensity as the response and an interaction term between driver type and database as the predictor. Driver intensity values for each driver were standardized between zero and one to make them comparable. Database represented a four-level categorical variable (Random global sampling, Living Planet, BioTIME or PREDICTS database; in the marine model the PREDICTS database was omitted since it only covers the terrestrial realm). The 'Random global sampling' level was used as the reference so the coefficients for the three databases represent differences from the random global sampling. We fitted our model using the package brms v.2.15.0[97] and the default weakly informative priors. We considered credible intervals around the effect size (posterior mean) that do not overlap zero to indicate that global change on sites with existing biodiversity data differs from random sampling.

When effect sizes are negative this indicates that sites with existing biodiversity data underestimate driver intensity and when effect sizes are positive this indicates that sites with existing biodiversity data overestimate driver intensity.

Quantifying mismatches between peak driver intensity and ecological monitoring

To quantify how well biodiversity captured variation in global change over time, we focused on changes in primary forest cover derived from the LUH database[47] and in temperature, derived from the CRU TS v.4.05 database[53] for the terrestrial realm and from the NOAA Extended Reconstructed SST v5 database[54] for the marine realm. We chose these focal drivers because they explain large amounts of the variation in global change in the terrestrial realm[3] and they have long-enough temporal data to allow us to determine the trajectory of change and assess its match with the timing of biodiversity data collection.

We visualized primary forest cover from the year 800 to 2014 for the location of each terrestrial site in the Living Planet and BioTIME databases and indicated when the monitoring began at each site. We were unable to complete a similar analysis for the marine realm because there are no available temporal data for human use drivers like fishing of a sufficiently high temporal and spatial resolution. We extracted monthly mean temperature data for the same locations and summarized it as yearly averages. We then compared the slopes of temperature change during the biodiversity monitoring with the slopes of temperature change in the period preceding the monitoring (the two comparison periods were of equal length and always more than five years). For the comparison, we used general linear models predicting temperature change as a function of period, a two-level categorical variable with the levels of before and during monitoring.

Determining geographic and ecoregion representation

We mapped the location of sampling sites within the Living Planet, BioTIME and PREDICTS databases. Ecoregion polygons were retrieved for the terrestrial[98] and marine[99] realms. We then counted the number of ecoregions that were sampled by each database (sampling indicates at least one record in a given ecoregion).

Determining taxonomic representation

To quantify taxonomic representation, we extracted the numbers of known species per taxa from the 2019 edition of the Catalogue of Life (http://www.catalogueoflife.org) and then compared them to the numbers of distinct species recorded in the Living Planet, BioTIME and PREDICTS databases. We quantified taxonomic representation as percentages of species which have at least one record in the respective databases.

Data availability

All data are publicly available. Population and biodiversity time-series data are freely available in the Living Planet and BioTIME Databases (see references for details on data collection). The Living Planet Database can be accessed on http://www.livingplanetindex.org/data_portal. The BioTIME Database can be accessed on Zenodo (https://doi.org/10.5281/zenodo.1211105) or through the BioTIME website (http://biotime.st-andrews.ac.uk/). PREDICTS can be downloaded from https://www.predicts.org.uk/pages/outputs.html. The database of biodiversity data locations and associated global change driver magnitudes we compiled is available on GitHub (https://github.com/gndaskalova/GlobalChangeSpace).

Code availability

The R code for data manipulation, analyses and data visualization is available on GitHub (https://github.com/gndaskalova/GlobalChangeSpace).

References

1. Ellis, E. C. *et al.* Used planet: A global history. *Proceedings of the National Academy of Sciences* **110**, 7978–7985 (2013).
2. IPCC. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp. (2014).
3. Bowler, D. E. *et al.* Mapping human pressures on biodiversity across the planet uncovers anthropogenic threat complexes. *People and Nature* **2**, 380–394 (2020).
4. IPBES. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (2019).
5. Culina, A. *et al.* Navigating the unfolding open data landscape in ecology and evolution. *Nat Ecol Evol* **2**, 420–426 (2018).
6. Hampton, S. E. *et al.* Big data and the future of ecology. *Frontiers in Ecology and the Environment* **11**, 156–162 (2013).
7. WWF. Living Planet Report - 2018: Aiming Higher. Grooten, M. and Almond, R.E.A.(Eds). WWF, Gland, Switzerland. (2018).
8. Dornelas, M. *et al.* BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography* **27**, 760–786 (2018).
9. Hudson, L. N. *et al.* The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and Evolution*

7, 145–188 (2017).

10. GBIF. GBIF Home Page. Available from: https://www.gbif.org [28 April 2021]. (2021).
11. Santini, L., Isaac, N. J. B. & Ficetola, G. F. TetraDENSITY: A database of population density estimates in terrestrial vertebrates. *Global Ecol Biogeogr* **27**, 787–791 (2018).
12. McCallen, E. *et al.* Trends in ecology: shifts in ecological research themes over the past four decades. *Front Ecol Environ* **17**, 109–116 (2019).
13. Daskalova, G. N. *et al.* Landscape-scale forest loss as a catalyst of population and biodiversity change. *Science* **368**, 1341–1347 (2020).
14. Antão, L. H. *et al.* Temperature-related biodiversity change across temperate marine and terrestrial systems. *Nat Ecol Evol* **4**, 927–933 (2020).
15. Millette, K. L. *et al.* No consistent effects of humans on animal genetic diversity worldwide. *Ecol Lett* **23**, 55–67 (2020).
16. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).
17. Vellend, M. *et al.* Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences* **110**, 19456–19459 (2013).
18. Dornelas, M. *et al.* Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. *Science* **344**, 296–299 (2014).
19. Blowes, S. A. *et al.* The geography of biodiversity change in marine and terrestrial

assemblages. *Science* **366**, 339–345 (2019).

20. Macgregor, C. J., Williams, J. H., Bell, J. R. & Thomas, C. D. Moth biomass increases and decreases over 50 years in Britain. *Nat Ecol Evol* **3**, 1645–1649 (2019).
21. Betts, M. G. *et al.* Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* **547**, 441–444 (2017).
22. Outhwaite, C. L., Gregory, R. D., Chandler, R. E., Collen, B. & Isaac, N. J. B. Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nat Ecol Evol* **4**, 384–392 (2020).
23. Daskalova, G. N., Myers-Smith, I. H. & Godlee, J. L. Rare and common vertebrates span a wide spectrum of population trends. *Nat Commun* **11**, 4394 (2020).
24. Gonzalez, A. *et al.* Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology* **97**, 1949–2960 (2016).
25. Amano, T., Lamming, J. D. L. & Sutherland, W. J. Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. *BioScience* **66**, 393–400 (2016).
26. Boakes, E. H. *et al.* Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biol* **8**, e1000385 (2010).
27. Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. Global priorities for an effective information basis of biodiversity distributions. *Nat Commun* **6**, 8221 (2015).
28. Hill, S. L. L. *et al. Worldwide impacts of past and projected future land-use change on local species richness and the Biodiversity Intactness Index*.

http://biorxiv.org/lookup/doi/10.1101/311787 (2018) doi:10.1101/311787.

29. Schipper, A. M. *et al.* Projecting terrestrial biodiversity intactness with GLOBIO 4. *Glob Change Biol* gcb.14848 (2019) doi:10.1111/gcb.14848.
30. OECD. "The Post-2020 Biodiversity Framework: Targets, indicators and measurability implications at global and national level", November version. (2019).
31. Xu, H. *et al.* Ensuring effective implementation of the post-2020 global biodiversity targets. *Nat Ecol Evol* **5**, 411–418 (2021).
32. Johnston, A., Moran, N., Musgrove, A., Fink, D. & Baillie, S. R. Estimating species distributions from spatially biased citizen science data. *Ecological Modelling* **422**, 108927 (2020).
33. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. Taxonomic bias in biodiversity data and societal preferences. *Sci Rep* **7**, 9132 (2017).
34. Mentges, A., Blowes, S. A., Hodapp, D., Hillebrand, H. & Chase, J. M. Effects of site-selection bias on estimates of biodiversity change. *Conservation Biology* **35**, 688–698 (2021).
35. Mihoub, J.-B. *et al.* Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. *Scientific Reports* **7**, 41591 (2017).
36. Wauchope, H. S. *et al.* Evaluating Impact Using Time-Series Data. *Trends in Ecology & Evolution* **36**, 196–205 (2021).
37. Buckland, S. T. & Johnston, A. Monitoring the biodiversity of regions: Key principles and possible pitfalls. *Biological Conservation* **214**, 23–34 (2017).

38. Feng, X. *et al.* A review of the heterogeneous landscape of biodiversity databases: opportunities and challenges for a synthesized biodiversity knowledge base. *EcoEvoArxiv preprint* (2021) doi:https://doi.org/10.32942/osf.io/9zt4a.

39. Shirey, V., Belitz, M. W., Barve, V. & Guralnick, R. A complete inventory of North American butterfly occurrence data: narrowing data gaps, but increasing bias. *Ecography* **44**, 537–547 (2021).

40. Marsh, D. M. & Cosentino, B. J. Causes and consequences of non-random drop-outs for citizen science projects: lessons from the North American amphibian monitoring program. *Freshwater Science* **38**, 292–302 (2019).

41. Ritchie, P. D. L., Clarke, J. J., Cox, P. M. & Huntingford, C. Overshooting tipping point thresholds in a changing climate. *Nature* **592**, 517–523 (2021).

42. Moore, J. C. Predicting tipping points in complex environmental systems. *Proc Natl Acad Sci USA* **115**, 635–636 (2018).

43. Isbell, F., Tilman, D., Reich, P. B. & Clark, A. T. Deficits of biodiversity and productivity linger a century after agricultural abandonment. *Nat Ecol Evol* **3**, 1533–1538 (2019).

44. De Palma, A. *et al.* Challenges With Inferring How Land-Use Affects Terrestrial Biodiversity: Study Design, Time, Space and Synthesis. in *Advances in Ecological Research* vol. 58 163–199 (Elsevier, 2018).

45. Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).

46. Osenberg, C. W., Schmitt, R. J., Holbrook, S. J., Abu-Saba, K. E. & Flegal, A. R.

Detection of Environmental Impacts: Natural Variability, Effect Size, and Power Analysis. *Ecological Applications* **4**, 16–30 (1994).

47. Hurtt, G. C. *et al.* Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change* **109**, 117–161 (2011).
48. Klein Goldewijk, K., Beusen, A., Doelman, J. & Stehfest, E. Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth Syst. Sci. Data* **9**, 927–953 (2017).
49. Dakos, V. *et al.* Ecosystem tipping points in an evolving world. *Nat Ecol Evol* **3**, 355–362 (2019).
50. Jarzyna, M. A. & Jetz, W. Taxonomic and functional diversity change is scale dependent. *Nature Communications* **9**, (2018).
51. Jönsson, G. M., Broad, G. R., Sumner, S. & Isaac, N. J. B. A century of social wasp occupancy trends from natural history collections: spatiotemporal resolutions have little effect on model performance. *Insect Conserv Divers* icad.12494 (2021) doi:10.1111/icad.12494.
52. Jennions, M. D. A survey of the statistical power of research in behavioral ecology and animal behavior. *Behavioral Ecology* **14**, 438–445 (2003).
53. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data* **7**, 109 (2020).
54. Huang, B. *et al.* Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate* **30**, 8179–8205 (2017).

55. Titley, M. A., Snaddon, J. L. & Turner, E. C. Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PLoS ONE* **12**, e0189577 (2017).

56. Hallmann, C. A. *et al.* More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE* **12**, e0185809 (2017).

57. Sánchez-Bayo, F. & Wyckhuys, K. A. G. Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation* **232**, 8–27 (2019).

58. van Klink, R. *et al.* Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* **368**, 417–420 (2020).

59. Dornelas, M. & Daskalova, G. N. Nuanced changes in insect abundance. *Science* **368**, 368–369 (2020).

60. Didham, R. K. *et al.* Interpreting insect declines: seven challenges and a way forward. *Insect Conserv Divers* **13**, 103–114 (2020).

61. Montgomery, G. A. *et al.* Is the insect apocalypse upon us? How to find out. *Biological Conservation* 108327 (2019) doi:10.1016/j.biocon.2019.108327.

62. Daskalova, G. N., Phillimore, A. B. & Myers-Smith, I. H. Accounting for year effects and sampling error in temporal analyses of invertebrate population and biodiversity change: a comment on Seibold *et al* . 2019. *Insect Conserv Divers* **14**, 149–154 (2021).

63. Olson, D. M. & Dinerstein, E. The Global 200: Priority Ecoregions for Global Conservation. *Annals of the Missouri Botanical Garden* **89**, 199–224 (2002).

64. Isaac, N. J. B. & Cowlishaw, G. How species respond to multiple extinction threats.

Proceedings of the Royal Society of London. Series B: Biological Sciences **271**, 1135–1141 (2004).

65. Rocha-Ortega, M., Rodriguez, P. & Córdoba-Aguilar, A. Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. *Ecol Entomol* een.13027 (2021) doi:10.1111/een.13027.
66. Barnagaud, J.-Y. *et al.* Relating Habitat and Climatic Niches in Birds. *PLoS ONE* **7**, e32819 (2012).
67. Frishkoff, L. O. *et al.* Climate change and habitat conversion favour the same species. *Ecology Letters* **19**, 1081–1090 (2016).
68. Klink, R. *et al.* InsectChange: a global database of temporal changes in insect and arachnid assemblages. *Ecology* **102**, (2021).
69. Brlík, V. *et al.* Long-term and large-scale multispecies dataset tracking population changes of common European breeding birds. *Sci Data* **8**, 21 (2021).
70. Pereira, H. M. *et al.* Essential Biodiversity Variables. *Science* **339**, 277–278 (2013).
71. Collen, B., Ram, M., Zamin, T. & McRae, L. The Tropical Biodiversity Data Gap: Addressing Disparity in Global Monitoring. *Tropical Conservation Science* **1**, 75–88 (2008).
72. Cameron, E. K. *et al.* Global gaps in soil biodiversity data. *Nat Ecol Evol* **2**, 1042–1043 (2018).
73. Geijzendorffer, I. R. *et al.* Bridging the gap between biodiversity data and policy reporting needs: An Essential Biodiversity Variables perspective. *J Appl Ecol* **53**, 1341–1350 (2016).

74. Wetzel, F. T. *et al.* Unlocking biodiversity data: Prioritization and filling the gaps in biodiversity observation data in Europe. *Biological Conservation* **221**, 78–85 (2018).

75. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).

76. Cameron, E. K. *et al.* Global mismatches in aboveground and belowground biodiversity. *Conservation Biology* **33**, 1187–1192 (2019).

77. Dornelas, M. *et al.* A balance of winners and losers in the Anthropocene. *Ecology Letters* **22**, 847–854 (2019).

78. Mace, G. M. *et al.* Aiming higher to bend the curve of biodiversity loss. *Nat Sustain* **1**, 448–451 (2018).

79. Leclère, D. *et al.* Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* (2020) doi:10.1038/s41586-020-2705-y.

80. Pereira, L. M. *et al.* Developing multiscale and integrative nature–people scenarios using the Nature Futures Framework. *People and Nature* **2**, 1172–1195 (2020).

81. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. How Many Species Are There on Earth and in the Ocean? *PLoS Biol* **9**, e1001127 (2011).

82. McRae, L., Deinet, S. & Freeman, R. The Diversity-Weighted Living Planet Index: Controlling for Taxonomic Bias in a Global Biodiversity Indicator. *PLoS ONE* **12**, e0169156 (2017).

83. Leung, B. *et al.* Clustered versus catastrophic global vertebrate declines. *Nature* **588**, 267–271 (2020).

84. Buschke, F. T., Hagan, J. G., Santini, L. & Coetzee, B. W. T. Random population

fluctuations bias the Living Planet Index. *Nat Ecol Evol* (2021) doi:10.1038/s41559-021-01494-0.

85. Palma, A. D. *et al. Annual changes in the Biodiversity Intactness Index in tropical and subtropical forest biomes, 2001-2012*. http://biorxiv.org/lookup/doi/10.1101/311688 (2018) doi:10.1101/311688.
86. Sánchez-Tójar, A., Moran, N. P., O'Dea, R. E., Reinhold, K. & Nakagawa, S. Illustrating the importance of meta-analysing variances alongside means in ecology and evolution. *J Evol Biol* **33**, 1216–1223 (2020).
87. Bennington, C. C. & Thayne, W. V. Use and Misuse of Mixed Model Analysis of Variance in Ecological Studies. *Ecology* **75**, 717–722 (1994).
88. Tessarolo, G., Lobo, J. M., Rangel, T. F. & Hortal, J. High uncertainty in the effects of data characteristics on the performance of species distribution models. *Ecological Indicators* **121**, 107147 (2021).
89. Wintle, B. A., McCARTHY, M. A., Volinsky, C. T. & Kavanagh, R. P. The Use of Bayesian Model Averaging to Better Represent Uncertainty in Ecological Models. *Conservation Biology* **17**, 1579–1590 (2003).
90. Fitzpatrick, M. C. & Hargrove, W. W. The projection of species distribution models and the problem of non-analog climate. *Biodivers Conserv* **18**, 2255–2261 (2009).
91. Pfeiffer, M., Kumar, D., Martens, C. & Scheiter, S. Climate change will cause non-analog vegetation states in Africa and commit vegetation to long-term change. *Biogeosciences* **17**, 5829–5847 (2020).
92. Mahony, C. R., MacKenzie, W. H. & Aitken, S. N. Novel climates: Trajectories of

climate change beyond the boundaries of British Columbia's forest management knowledge system. *Forest Ecology and Management* **410**, 35–47 (2018).

93. Zurell, D., Elith, J. & Schröder, B. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions: Predicting to new environments. *Diversity and Distributions* **18**, 628–634 (2012).
94. Kissling, W. D. *et al.* Towards global data products of Essential Biodiversity Variables on species traits. *Nat Ecol Evol* **2**, 1531–1540 (2018).
95. Jetz, W. *et al.* Essential biodiversity variables for mapping and monitoring species populations. *Nat Ecol Evol* **3**, 539–551 (2019).
96. Jackson, A. L., Inger, R., Parnell, A. C. & Bearhop, S. Comparing isotopic niche widths among and within communities: SIBER - Stable Isotope Bayesian Ellipses in R: Bayesian isotopic niche metrics. *Journal of Animal Ecology* **80**, 595–602 (2011).
97. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* **80**, (2017).
98. Olson, D. M. *et al.* Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **51**, 933 (2001).
99. Spalding, M. D. *et al.* Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *BioScience* **57**, 573–583 (2007).

Acknowledgements: We thank the compilers and maintainers of the three open-access databases we used in our analyses – the Living Planet Database, BioTIME and PREDICTS.

Funding: G.N.D. was funded by a Carnegie-Caledonian PhD Scholarship and supported by a NERC doctoral training partnership grant (NE/L002558/1). M.D. was supported by a Leverhulme Trust Research Centre - the Leverhulme Centre for Anthropocene Biodiversity, by a Leverhulme Trust grant RPG-2019-401, by Natural Environment Research Council grants NE/V009338/1 and NE/T004487/1 and by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101003553. We thank the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig and the sChange working group for supporting the initial data synthesis work that has led to this study.

Competing interests: The authors declare no conflicts of interest.

Author contributions: GND conceptualized the study, conducted the analyses, visualization and investigation and wrote the first draft. DB, IMS and MD provided input on study conceptualization, analyses, investigation and methodology. DB, IMS and MD reviewed the manuscript drafts and all authors contributed to editing. IMS was the primary supervisor to GND, MD was the co-supervisor. DB compiled the global change driver layers (Bowler et al. 2020). GND acquired funding for a NERC Professional Internship Placement to collaborate with DB.