White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/212057/

Version: Published Version

Article:
Johnson, T.F., Beckerman, A.P., Childs, D.Z. et al. (2024) Revealing uncertainty in the status of biodiversity change. Nature, 628. pp. 788-794. ISSN: 0028-0836

https://doi.org/10.1038/s41586-024-07236-z

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

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

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Article

Revealing uncertainty in the status of biodiversity change

https://doi.org/10.1038/s41586-024-07236-z

Received: 23 November 2022

Accepted: 26 February 2024

Published online: 27 March 2024

Open access

Check for updates

T. F. Johnson[1]✉, A. P. Beckerman[1], D. Z. Childs[1], T. J. Webb[1], K. L. Evans[1], C. A. Griffiths[1,10], P. Capdevila[2,3,4], C. F. Clements[2], M. Besson[2,11], R. D. Gregory[5,6], G. H. Thomas[1], E. Delmas[1,7,8] & R. P. Freckleton[1,9]

Biodiversity faces unprecedented threats from rapid global change[1]. Signals of biodiversity change come from time-series abundance datasets for thousands of species over large geographic and temporal scales. Analyses of these biodiversity datasets have pointed to varied trends in abundance, including increases and decreases. However, these analyses have not fully accounted for spatial, temporal and phylogenetic structures in the data. Here, using a new statistical framework, we show across ten high-profile biodiversity datasets[2–11] that increases and decreases under existing approaches vanish once spatial, temporal and phylogenetic structures are accounted for. This is a consequence of existing approaches severely underestimating trend uncertainty and sometimes misestimating the trend direction. Under our revised average abundance trends that appropriately recognize uncertainty, we failed to observe a single increasing or decreasing trend at 95% credible intervals in our ten datasets. This emphasizes how little is known about biodiversity change across vast spatial and taxonomic scales. Despite this uncertainty at vast scales, we reveal improved local-scale prediction accuracy by accounting for spatial, temporal and phylogenetic structures. Improved prediction offers hope of estimating biodiversity change at policy-relevant scales, guiding adaptive conservation responses.

Accelerating rates of species extinction are driving global changes in biodiversity, threatening ecosystems and the services they provide[1]. In an attempt to reverse biodiversity declines, world leaders, policymakers and academics have called for action[12]. Evidence-based actions require long-term datasets and rigorous modelling to reliably detect and attribute biodiversity change through time[13,14]. At present, some of the most influential estimates of biodiversity change are calculated using datasets such as BioTIME[2], the Living Planet[15] or the North American Breeding Bird Survey[3]. Inferences from these abundance datasets have shaped policy[16] and are considered by some to be a key pillar of global biodiversity monitoring[17].

Biodiversity datasets are complex and typically subject to one or more sources of non-independence across the axes of time, space and evolution. This presents a challenge for analysis, as omission of even one of these sources of non-independence from a statistical model can lead to underestimation of uncertainty, incorrect trends and poorly resolved prediction, and ultimately undermines current interpretation of wildlife abundance trends[18–20]. A unifying feature of previous studies is that they are characterized by the consistent omission of one or more of these dependencies from their analysis. This imposes a risk that past estimates of abundance change—pointing to declines[15,21], no net change[18,22,23] and recovery[24]—may be unreliable.

Non-independence can be classified in a variety of ways, which we split into two core types: hierarchical, for which observations are pseudoreplicated or nested (for example, multiple trends for a given species, site or region in time); and correlative, for which observations become increasingly correlated (sometimes termed autocorrelation) when close in time[25], space[26] or phylogeny[27]. Under correlative non-independence, we may expect sequential abundance values in a time series to be more similar, and trends should be similar when near in space or in closely related species (Fig. 1). Although studies commonly account for hierarchical non-independence using features such as random effects in mixed models, a literature review covering hundreds of papers published in high-impact journals since 2010 revealed that studies rarely account for correlative non-independence across space (accounted for in 7% of studies), phylogeny (14%) or time (32%; Supplementary Table 1). Further, no biodiversity model has yet been formalized to account for all three sources of correlative non-independence at the same time.

Here we show that ignoring non-independence has serious consequences for inference of biodiversity trends. We introduce the correlated effect model, which incorporates hierarchical non-independence and all three sources of correlative non-independence, and apply it to ten high-profile, multi-species datasets that have been used to infer

[1]School of Biosciences, Ecology and Evolutionary Biology, University of Sheffield, Sheffield, UK. [2]School of Biological Sciences, Biosciences, University of Bristol, Bristol, UK. [3]Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Universitat de Barcelona (UB), Barcelona, Spain. [4]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain. [5]RSPB Centre for Conservation Science, The Lodge, Sandy, UK. [6]Centre for Biodiversity & Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK. [7]Habitat, Montreal, Quebec, Canada. [8]Institut des Sciences de la Forêt Tempérée, Université du Québec en Outaouais, Ripon, Quebec, Canada. [9]Debrecen Biodiversity Centre, University of Debrecen, Debrecen, Hungary. [10]Present address: Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of Marine Research, Lysekil, Sweden. [11]Present address: Sorbonne Université, CNRS, Biologie Intégrative des Organismes Marins, BIOM, Banyuls-sur-Mer, France. ✉e-mail: t.f.johnson@sheffield.ac.uk

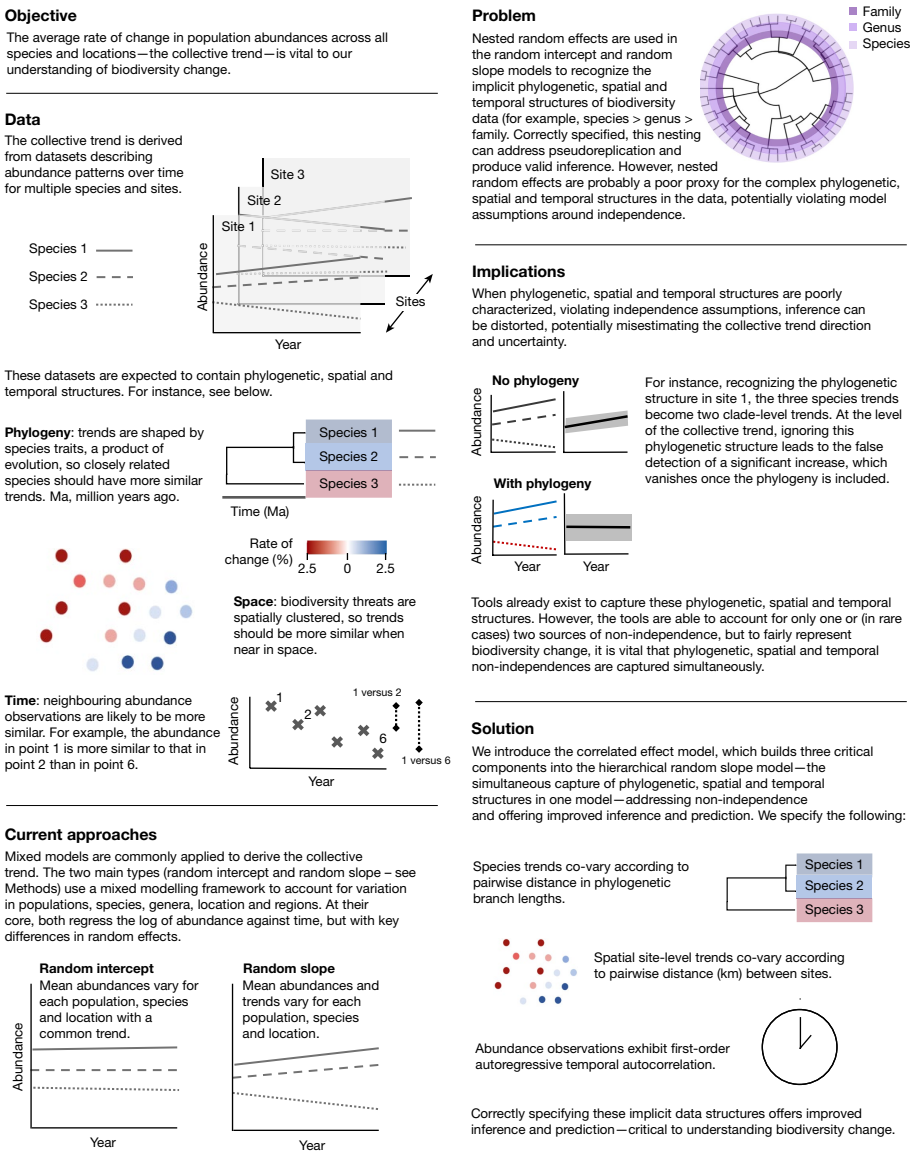


Fig. 1 | Impact of correlative non-independence on collective abundance trends. The text and images show the objective, implicit and key features of large-scale abundance datasets, current approaches for analysis, the problem, its implications and the solution.

abundance trends in global biodiversity[2–11]. Combined, these datasets describe the abundance (including relative abundance and densities) patterns of more than 30,000 populations, representing about 3,100 species and about 6,000 unique locations, and are considered some of the best biodiversity monitoring datasets available.

Non-independence increases uncertainty

We compared our correlated effect model with two mixed-effect modelling frameworks that are commonly used and account only for hierarchical non-independence: random intercept and random slope (both described in Fig. 1). Across the 44 relevant studies identified in a literature search spanning 282 published papers, 43% ($n = 19$) used a version of the random intercept model and 50% ($n = 22$) used a version of the random slope model (Supplementary Table 1). Comparing these commonly applied approaches to the correlated effect model, we detect a pronounced shift in collective abundance trends (that is, the model-derived average rate of change in abundance across all species and locations), and show that existing approaches underestimate collective trend uncertainty and can misestimate direction (Fig. 2).

Collective abundance trend uncertainty (that is, the standard deviation (s.d.) around the abundance–time coefficient) was underestimated in all ten datasets in both the random intercept and random slope models. These underestimates are large, with uncertainty in the correlated effect model 26 times greater [95% confidence interval (CI): 14–47] than that in the random intercept model and 3.4 times greater [95% CI: 1.8–6.2] than that in the random slope model. Further, after accounting for correlative non-independence, we find instances in which the trend direction shifts and even reverses (for example, from negative to positive). For instance, in the Living Planet dataset, a decreasing trend in the random intercept model shifts to a stable trend in the random slope model, before shifting back to a sharp albeit uncertain decrease after accounting for correlative non-independence. In three databases—the Living Planet, RivFishTIME and Atlantic reef fishes—the mean trends were more extreme under the correlated effect model, shifting away from zero (that is, no net change in abundance), although still highly uncertain. Across the three models, we observed complete agreement in trend direction and significance status (50% credible intervals) in only four of the datasets. At 95% credible intervals, we found no instances in which models agreed on trend direction and significance status.

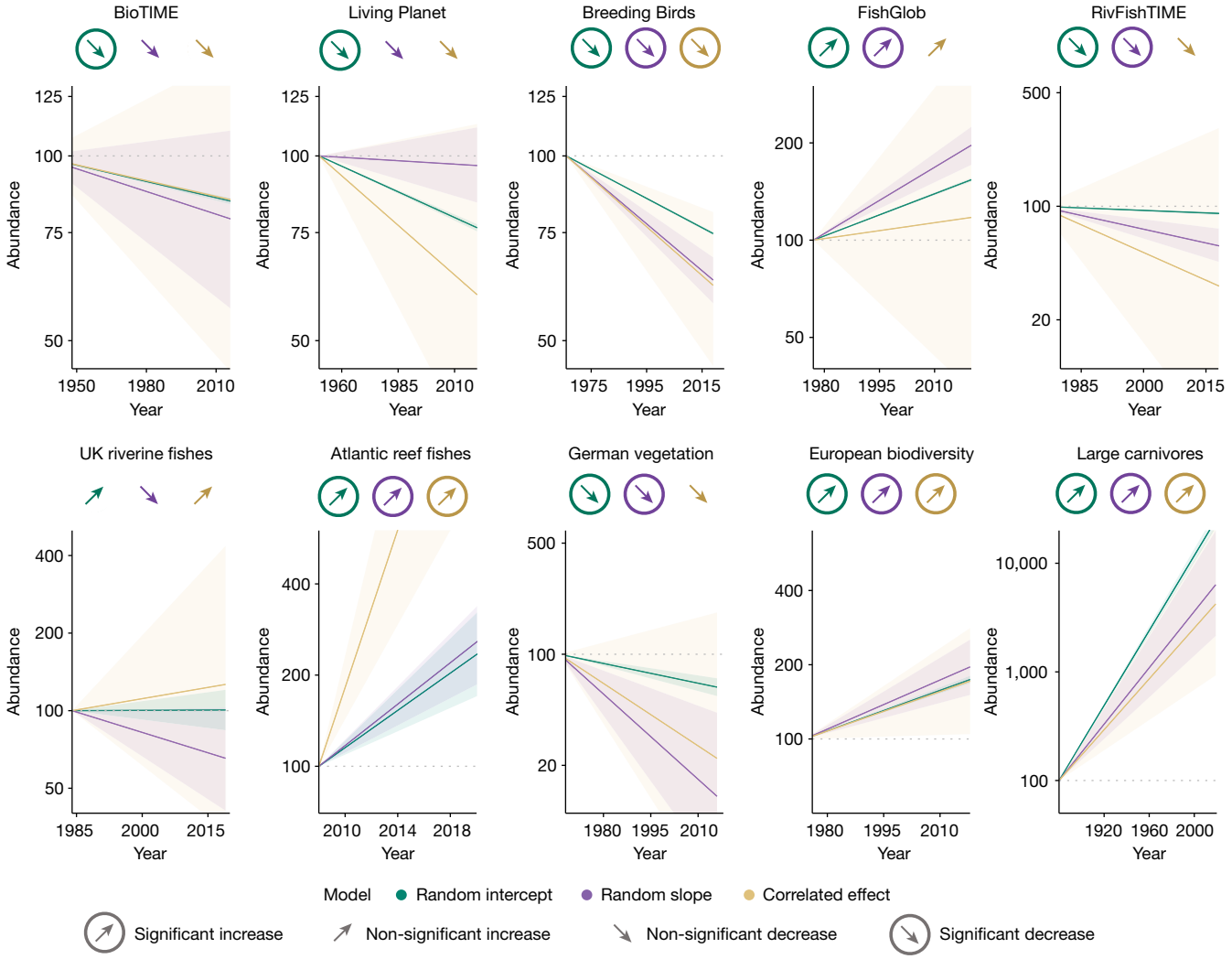


Fig. 2 | Widely used statistical models misrepresent biodiversity abundance trends. Abundance trend projections across ten high-profile datasets under three different models. Circles represent the collective trend (the coefficient describing the change in abundance over time averaged across all species and locations) for each dataset in our three models (from left to right): random intercept, random slope and the correlated effect model that simultaneously accounts for temporal, spatial and phylogenetic correlative non-independence. We specify four categories of trend: significant increase—coefficient is positive and significant; non-significant increase—coefficient is positive but not significant (that is, no detectable change); non-significant decrease—coefficient is negative but not significant (that is, no detectable change); significant decrease—coefficient is negative and significant. Significance indicates that the coefficient does not overlap zero at a 50% credible interval. Coefficients and 95% credible intervals are available in Supplementary Table 4. We use the collective trend coefficient and 50% credible intervals (represented by shading) to produce abundance projections for each model in each dataset from an arbitrary baseline abundance of 100. Abundance projections cover the time span of the observed data and are presented on the $\log_{10}$ scale.

Collective abundance trend uncertainty is likely to be underestimated when hierarchical terms (for example, random effects) fail to effectively represent the complex spatial, phylogenetic and temporal structures in the data (Extended Data Fig. 1). This is an apparently common phenomenon given all ten datasets underestimate uncertainty, and across the ten datasets, we find that correlative terms proportionally account for approximately one-third of the variation in the data (spatial: mean = 0.34 s.d. = 0.3; phylogeny: mean = 0.41, s.d. = 0.28), relative to the combined variance captured by the respective hierarchical and correlated terms. There is no comparable metric for the temporal term that describes the correlation between abundances instead of covariance between trends. Notably, the stark increase in uncertainty is not a consequence of simply introducing additional correlated terms. This is because uncertainty tends to increase substantially only when the correlated terms are capturing a high proportion of variance ($\beta = 1.00$, 95% CI: −0.19 to 2.21, $P = 0.09$; Extended Data Fig. 1). Through iteratively introducing the correlated terms into the random slope model (exploring six further model structures), it is apparent that uncertainty increases most after the inclusion of spatial correlation (Extended Data Fig. 2).

Predicting biodiversity change

Counterintuitively, accounting for correlative non-independence improves our capacity to make predictions 'out of sample'—that is, for a withheld subset of data not used to develop the model—despite the large uncertainty at the level of the collective trend. Part of the value of these abundance trends is that they can be used to estimate which species and locations are likely to be declining or recovering, and when. To evaluate whether the correlated effect model improves our ability to make local-scale predictions, we tested each model's ability to forecast new abundance observations and estimate population

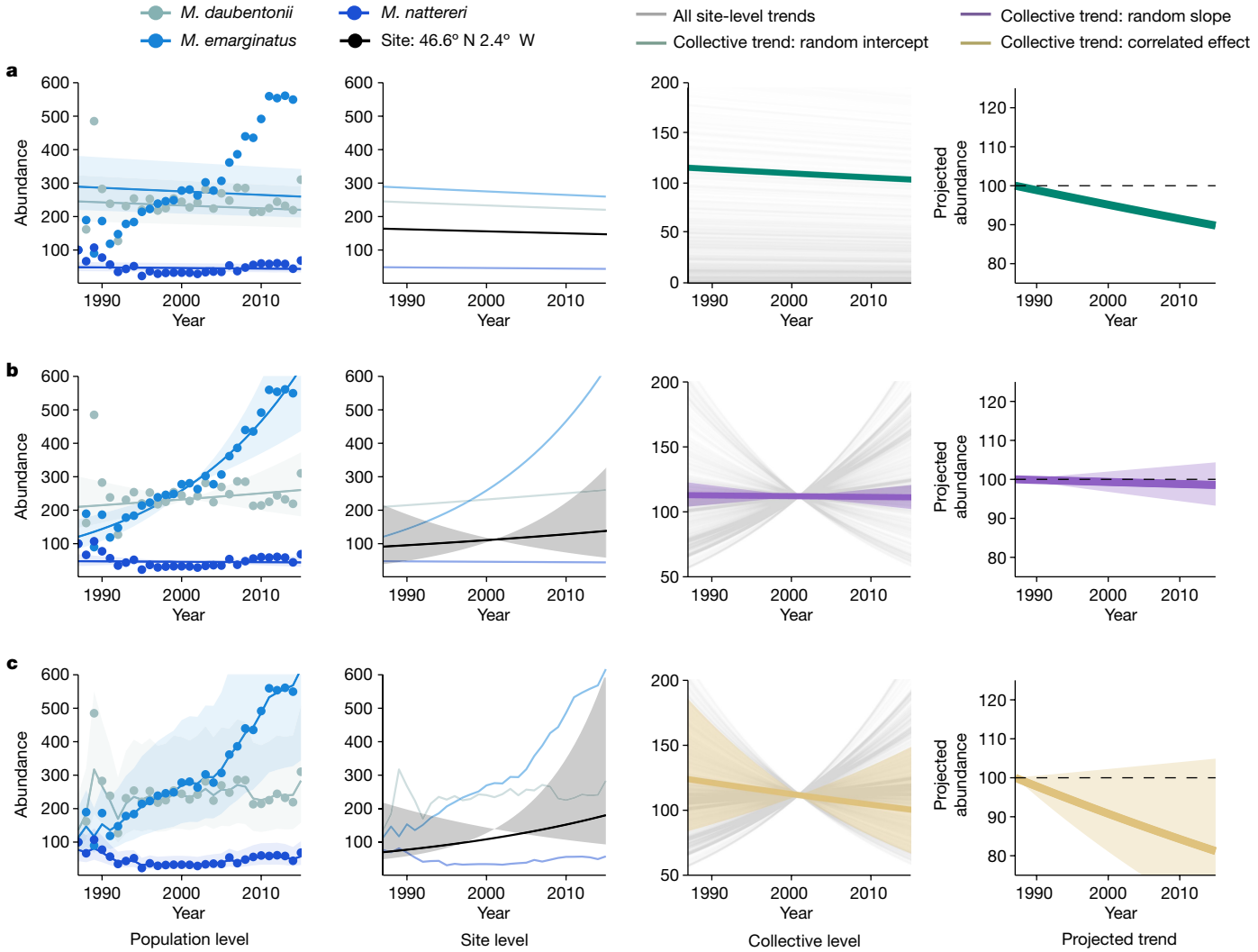


Fig. 3 | More complex models better represent population dynamics and improve the validity of conclusions across ecological scales. a–**c**, Example of how the three models (random intercept (**a**), random slope (**b**) and correlated effect (**c**)) describe abundance patterns at different ecological scales (finer ecological scales on the left). The population-level column showcases how each of the three models produce different estimates of abundance trends (lines are the median values with 95% credible interval shading) for all three bat species (genus *Myotis*) with data in a given location, with data points representing the observed abundance values. The site-level column depicts how the species' trends, under different models, influence the site-level trend (that is, a trend for a given location; black), in which the line and 95% credible intervals describe the median trend and variability in trend (respectively) across all species in the given location. At the collective level, the median trend for each unique site is represented by a faded grey line, and the median collective trend coefficient and 95% credible intervals are depicted by the coloured line and shading. At the site and collective levels, credible intervals solely describe uncertainty in the main parameter of interest, the rate of change coefficient, not the intercept. The final column describes how a hypothetical population would change under the median collective trend coefficient and 50% credible intervals projected from a relative baseline abundance of 100. This example is based on data in the Living Planet. In each plot, we restrict the time frame to the temporal extent of the population-level trends (1987–2015), instead of the total temporal extent of our Living Planet sample.

trends. For each dataset, we removed the final abundance observation in 50% of the population abundance time series and then evaluated each of the three models' ability to predict this value. Next, we conducted leave-one-out cross-validation to assess trend prediction, removing a population time series (that is, trend) from each dataset and testing each model's ability to recover this population's abundance trend. We repeated this cross-validation 50 times for each of the 10 datasets. In each dataset, we report predictive accuracy for each of these approaches as the percentage error (PE), a metric describing the median of the absolute percentage difference between predicted and observed values; for example, with a 5% error, an abundance on the log scale of 1 would become 1.05. Summarizing across datasets, we report the mean and s.d.

Across the 10 datasets, the correlated effect model estimated the final abundance observation with 16.1% error (s.d. = 7.5%), 1.51 times more accurately than the random intercept model (mean = 24.4%, s.d. = 16.2%) and 1.13 times more accurately than the random slope model (mean = 18.3%; s.d. = 10.5%). The correlated effect model also performed best when estimating missing population trends, with an error of 18.3% (s.d. = 11.6%), 1.35 times more accurate than the random slope model (mean = 28.9%; s.d. = 25.5%). In one case, using the correlated effect model to capture the spatial, temporal and phylogenetic structures halved the trend prediction error, relative to the random slope model. The random slope model had a lower prediction error than the correlated effect model in only one dataset in the abundance assessment, and two datasets in the trend assessment.

The improved prediction in the correlated effect model is a consequence of handling temporal, spatial and phylogenetic non-independence. For the Living Planet data, recognizing temporal correlation between sequential abundance values ($\rho = 0.52$) introduces nonlinearity (residual variability in linear trends) in temporal trends, and more closely represents realistic population dynamics (Fig. 3, population level). Comparably, temporal non-independence in the Living Planet data was higher than the average across the ten datasets ($\rho = 0.31$, s.d. = 0.42, range: −0.65 to 0.99). Accounting for this temporal structure in population trends can also influence trend direction and uncertainty of species-level and site-level trends, relative to the random slope model (Fig. 3, site level). At the global level, the presence of temporal, spatial (proportion of variance captured by spatial correlation term = 0.30) and phylogenetic (proportion of variance captured by phylogenetic correlation term = 0.34) structures elevated the uncertainty around the overall trend (Extended Data Fig. 2), ultimately leading to more robust inference.

The presence of the observed spatial and phylogenetic structures also has the added benefit of allowing us to make predictions beyond the species and location data (Fig. 4), offering important insight into species and locations most likely to exhibit declines and recoveries. Our ability to predict a given species trend is dependent on the species being contained and accurately described in a phylogeny. Efforts continue to expand the breadth and quality of phylogenetic information[28], and across our 10 datasets, we were able to obtain phylogenetic information for 80% of species.

Implications for biodiversity science

The abundance datasets we analyse are influential in policy, tracking progress towards biodiversity targets at national and international scales[16,17,29], and so it is vital that any inference gained is both valid and reliable. Our work shows that when biodiversity change datasets are analysed without accounting for correlative non-independence among phylogeny, time and space, there is a substantial risk that trend uncertainty could be underestimated, trend direction misestimated and policy misinformed. Further, once uncertainty is appropriately attributed in the correlated effect model, we failed to detect a single significant trend in collective abundance across the 10 datasets under 95% credible intervals. This pervasive pattern points to a highly uncertain status in collective abundance trends; that is, it is unclear how biodiversity is changing across vast spatial and taxonomic scales once uncertainty is appropriately estimated.

The random intercept model, used by 43% of studies, underestimated trend uncertainty 26-fold. The random slope model, used by 50% of studies, performs better but still underestimates uncertainty 3.4-fold. This underestimated uncertainty has a substantial impact on trend inference, for which 9 datasets have significant trends at 50% credible intervals in the random intercept model compared to 7 datasets in the random slope model, and just 4 datasets in the correlated effect model. At 95% credible intervals, we found 8 significant collective trends in the random intercept model, 4 significant trends in the random slope model and zero significant trends in the correlated effect model. This raises questions about the robustness of existing estimates of abundance change in the literature.

Past estimates of abundance change have pointed to declines[15,21], no net change[18,22,23] and recovery[24]. This high variability across studies and datasets could be well founded given their different spatial, temporal and taxonomic scales, but it is also paradoxical given the expectation that biodiversity has declined under intense and widespread global change[1]. In the correlated effect model, we partially resolve this variability between datasets, as our results generalize under the common feature of substantial uncertainty. However, the absence of significant trends in the correlated effect model also further emphasizes the paradox of failing to detect biodiversity loss despite rapid global change. Ours is not the first study to fail to detect declining abundances. For instance, previous work has shown that most abundance trends exhibit no net change[23], and that the magnitude of decline reduces after accounting for extreme values[18] and random abundance fluctuations[19]. Other work suggests the current data collection infrastructure is too small and biased to detect a trend reliably[30]. Similarly, analyses of BioTIME suggest that declines are unlikely because environmental change generates winners as well as losers[2,22], whose opposing population trends may cancel each other out at global scales.

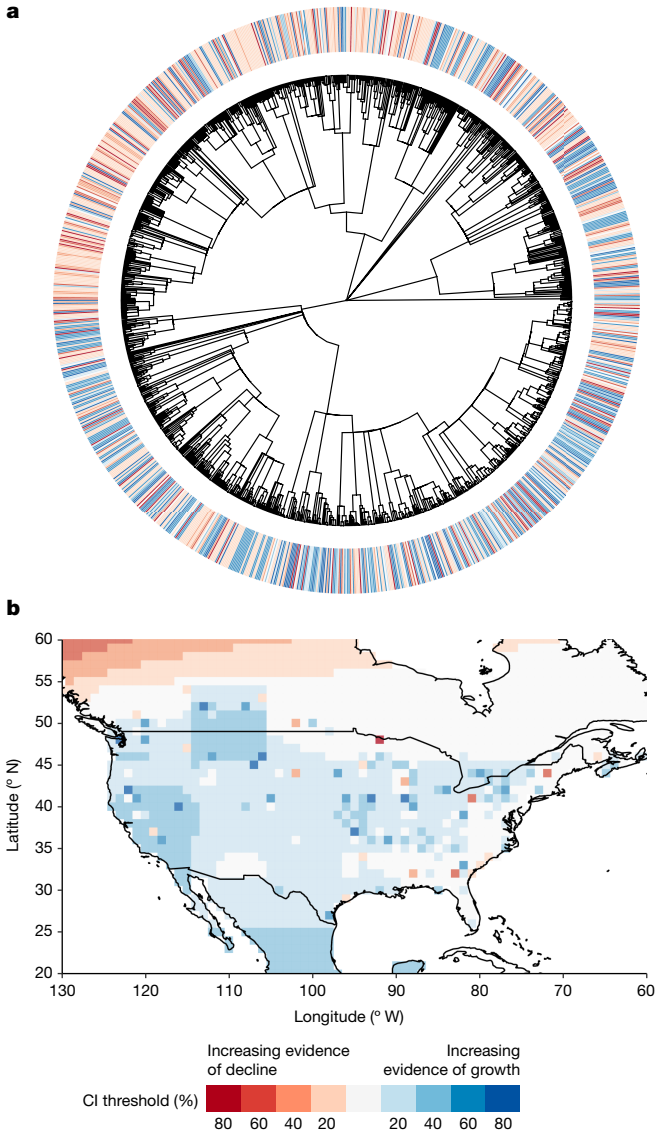


Fig. 4 | Abundance change varies over phylogenetic and spatial extents. Evidence of abundance change at different significance thresholds (for example, at an 80% CI threshold, dark red indicates evidence of declines whereas dark blue indicates evidence of increases). **a**, For the phylogeny, the species-level trends were derived by summing across hierarchical taxonomic random effects and phylogenetic correlation terms. Asymptotic species-level confidence thresholds were derived using uncertainty in phylogenetic predictions at multiple *z*-scores. To improve visualization, phylogenetic branch lengths are log transformed. **b**, For space, we take taxonomic and phylogenetic information from **a** for one iconic and abundant North American species, the American robin *Turdus migratorius*, and combine this with hierarchical and correlative spatial terms to make population-level predictions across terrestrial space. Asymptotic confidence thresholds were derived at the population scale (for example, species in a given site) using multiple *z*-scores. These predictions are drawn from the correlated effect model and BioTIME data (Supplementary Table 1).

All things considered, it seems likely that collective abundance trends over varied taxa exhibiting varied responses to varied degrees of environmental change may not present as significant even with vast quantities of data. Perhaps a more considered approach is necessary, focused on describing which taxonomic groups and specific locations are declining and recovering, and why. The correlated effect model is particularly well placed for exploring this question, as the integration of space and phylogeny allows us to explore the particular locations and clades for which abundance trends shift from stable to decline. Although the high uncertainty around collective trends limits our general understanding of abundance changes, we observe an increase in accuracy of abundance forecasts and trend predictions under the correlated effect model, delivering a much-needed improvement in prediction at local scales (Fig. 4). The more complex representation of space, time and phylogeny is key to this improved prediction, for which, as an example, a priori information on evolutionary history can help predict which species are likely to decline and recover. Our new methods offer the hope of greater clarity in biodiversity trend estimation across different datasets and geographies, to inform and guide adaptive conservation policy responses.

Despite failing to detect a decline in collective abundance across the ten datasets at 95% credible intervals, our results do not necessarily mean that wildlife abundances have not declined, or that the current estimates of trends are incorrect. It is possible that abundances may have increased on average, or perhaps declines have been far more extreme than we have previously imagined; simply, the uncertainty is too high to know. With this in mind, we re-emphasize calls to urgently expand data collection and improve trend detectability, but this will only help to estimate current and future biodiversity trends. Given the large uncertainty associated with our estimates of abundance change, it seems that past abundance patterns are lost and undetectable at present. A shift towards causal frameworks of detection and attribution is probably necessary to estimate past biodiversity change[14].

In our study, we have solely focused on deriving the collective abundance trend given its potential political importance[17], but the core statistical framework could be applied to other biodiversity data types (for example, occupancy data), adapted to other metrics (for example, species richness) and integrated into a global biodiversity observation system[14]. For instance, the species trend coefficients could be extracted from the model and used alongside estimates of absolute species abundance to determine the absolute change in populations[31]. Weightings could be included to increase the influence of common species, allowing us to reconcile and test for differences between collective abundance trends, biomass decline and individuals lost. Precise and accurate estimates of abundance change in time and space also underpin a variety of policy-relevant facets of biodiversity, including ecosystem function through abundance-weighted functional diversity[32,33], energy flux, and population stability and resilience[34].

The implications of our findings extend beyond biodiversity change. Spatial, temporal and phylogenetic data structures are common in ecological and evolutionary research. Under the presence of correlative non-independence, there is a potentially pervasive risk in the field that we have mis-specified our statistical models, violating data independence assumptions, and producing unreliable inference. However, this also presents an opportunity, as the correlative effect model could be adapted for a wide array of settings, improving inference across ecology and evolution. The model could also streamline workflows by simultaneously capturing spatial, phylogenetic and temporal structures, avoiding the need to capture terms in multiple separate analyses.

Our analytical advance offers new potential in predictive ecology, but given the severity of the potential implications of biodiversity loss[35], it is vital that we continue to expand and improve these methods. We offer a general framework for addressing spatial, temporal and phylogenetic non-independence, but further advances are necessary, considering underlying issues around time-series length[36], bias and non-probability[37], nonlinearity and varied responses to environmental change[38], modern data collection philosophies[39] and rigorous analysis approaches[40]. This combination of improved methods and data has the potential to reveal patterns of biodiversity change and disentangle the complex processes shaping our ecosystems.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07236-z.

1. *Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (IPBES, 2019).
2. Dornelas, M. et al. BioTIME: a database of biodiversity time series for the Anthropocene. *Glob. Ecol. Biogeogr.* **27**, 760–786 (2018).
3. Sauer, J. R. et al. The first 50 years of the North American Breeding Bird Survey. *Condor* **119**, 576–593 (2017).
4. Pilotto, F. et al. Meta-analysis of multidecadal biodiversity trends in Europe. *Nat. Commun.* **11**, 3486 (2020).
5. Comte, L. et al. RivFishTIME: a global database of fish time-series to study global change ecology in riverine systems. *Glob. Ecol. Biogeogr.* **30**, 38–50 (2021).
6. *Freshwater Fish Counts* (UK Government, 2021, accessed 1 May 2023); https://environment.data.gov.uk/dataset/ce2618db-d507-4671-bafe-840b930d2297.
7. Johnson, T. F., Cruz, P., Isaac, N. J. B., Paviolo, A. & González-Suárez, M. CaPTrends: a database of large carnivoran population trends from around the world. *Glob. Ecol. Biogeogr.* **31**, 2475–2482 (2022).
8. *Living Planet Index Database* (LPI, 2023, accessed 1 May 2023); https://www.livingplanetindex.org/data_portal.
9. Maureaud, A.A., Palacios-Abrantes, J., Kitchel, Z. et al. FISHGLOB_data: an integrated dataset of fish biodiversity sampled with scientific bottom-trawl surveys. *Sci. Data* **11**, 24 (2024).
10. Quimbayo, J. P. et al. TimeFISH: Long-term assessment of reef fish assemblages in a transition zone in the Southwestern Atlantic. *Ecology* **104**, e3966 (2023).
11. Jandt, U. et al. More losses than gains during one century of plant biodiversity change in Germany. *Nature* **611**, 512–518 (2022).
12. Ripple, W. J. et al. World scientists' warning to humanity: a second notice. *BioScience* **67**, 1026–1028 (2017).
13. Pereira, H. M. et al. Essential biodiversity variables. *Science* **339**, 277–278 (2013).
14. Gonzalez, A., Chase, J. M. & O'Connor, M. I. A framework for the detection and attribution of biodiversity change. *Philos. Trans. R. Soc. B* **378**, 20220182 (2023).
15. *Living Planet Report 2020 - Bending the Curve of Biodiversity Loss* (WWF, 2020).
16. Ledger, S. E. H. et al. Past, present, and future of the Living Planet Index. *npj Biodivers.* **2**, 12 (2023).
17. Geldmann, J., Byaruhanga, A., Gregory, R., Visconti, P. & Xu, H. Prioritize wild species abundance indicators. *Science* **380**, 591–592 (2023).
18. Leung, B. et al. Clustered versus catastrophic global vertebrate declines. *Nature* **588**, 267–271 (2020).
19. Buschke, F. T., Hagan, J. G., Santini, L. & Coetzee, B. W. T. Random population fluctuations bias the Living Planet Index. *Nat. Ecol. Evol.* **5**, 1145–1152 (2021).
20. Loreau, M. et al. Do not downplay biodiversity loss. *Nature* **601**, E27–E28 (2022).
21. van Klink, R. et al. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* **368**, 417–420 (2020).
22. Dornelas, M. et al. A balance of winners and losers in the Anthropocene. *Ecol. Lett.* **22**, 847–854 (2019).
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. Haase, P. et al. The recovery of European freshwater biodiversity has come to a halt. *Nature* **620**, 582–588 (2023).
25. Dornelas, M. et al. Quantifying temporal change in biodiversity: challenges and opportunities. *Proc. R. Soc. B* **280**, 20121931 (2013).
26. Browning, E., Freeman, R., Boughey, K. L., Isaac, N. J. B. & Jones, K. E. Accounting for spatial autocorrelation and environment are important to derive robust bat population trends from citizen science data. *Ecol. Indic.* **136**, 108719 (2022).
27. Dinnage, R., Skeels, A. & Cardillo, M. Spatiophylogenetic modelling of extinction risk reveals evolutionary distinctiveness and brief flowering period as threats in a hotspot plant genus. *Proc. R. Soc. B* **287**, 20192817 (2020).
28. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
29. *Status of Priority Species: Relative Abundance* (Department for Environment, Food & Rural Affairs, 2023).
30. Valdez, J. W. et al. The undetectability of global biodiversity trends using local species richness. *Ecography* **2023**, e06604 (2023).
31. Rosenberg, K. V. et al. Decline of the North American avifauna. *Science* **366**, 120–124 (2019).
32. Sol, D. et al. The worldwide impact of urbanisation on avian functional diversity. *Ecol. Lett.* **23**, 962–972 (2020).

33. Oliver, T. H. et al. Declining resilience of ecosystem functions under biodiversity loss. *Nat. Commun.* **6**, 10122–10122 (2015).
34. Capdevila, P., Noviello, N., McRae, L., Freeman, R. & Clements, C. F. Global patterns of resilience decline in vertebrate populations. *Ecol. Lett.* **25**, 240–251 (2022).
35. Rockström, J. et al. A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
36. White, E. R. Minimum time required to detect population trends: the need for long-term monitoring programs. *BioScience* **69**, 40–46 (2019).
37. Boyd, R. J., Powney, G. D. & Pescott, O. L. We need to talk about nonprobability samples. *Trends Ecol. Evol.* **38**, 521–531 (2023).
38. Elmqvist, T. et al. Response diversity, ecosystem change, and resilience. *Front. Ecol. Environ.* **1**, 488–494 (2003).
39. Meng, X.-L. Statistical paradises and paradoxes in big data (I): low of large populations, big data paradox, and the 2016 US presidential election. *Ann. Appl. Stat.* **12**, 685–726 (2018).
40. Wauchope, H. S. et al. Protected areas have a mixed impact on waterbirds, but management helps. *Nature* **605**, 103–107 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024

Methods

Data

We compiled ten datasets that describe population abundances through time[2–11,41]. These datasets represent some of the most influential in ecology and conservation biology, forming a basis for influential reports such as the Living Planet[15], as well as a series of high-profile and highly cited publications (see Supplementary Table 2 for a full summary). For each dataset, we extracted the population abundance estimates, the accompanying time-stamps, the scientific names of the species, the name of the site (location) where the population was sampled and any site coordinates. For datasets to be included, they had to be open access, and contain multiple abundance time series for a selection of species and locations. Although these datasets are vital in biodiversity science, many of the datasets are prone to biases (for example, lacking tropical representation, and contain few plant and invertebrate species). The datasets have been compiled from a variety of methods, realms and systems, covering a vast array of spatial, taxonomic and temporal scales. Further, there is probably some overlap in data between datasets where population time series may occur in more than one dataset. We take no action to correct or acknowledge these biases and features, as our analysis is designed to show how model choice can have a substantial influence on inference in a variety of datasets, rather than to derive a consensus trend across datasets.

To account for correlative non-independence introduced by species' shared evolutionary past, we extracted a phylogeny for each dataset. We used synthetic trees from the Open Tree of Life[42,43] and estimated missing branch lengths using Grafen's approach[44] from the compute.brlen function in the R package ape[45]. The Open Tree of Life identified a phylogeny for 80% of species (n = 23,871); all other species were removed from the analysis. For studies with the overall aim of assessing biodiversity change, removing species could be problematic, as the collective trend would not be representative of all species. However, in our case, in which the aim is to assess how collective trend inference changes under a variety of modelling approaches, trimming the data to species with an accompanying phylogeny has no impact on our conclusions. Regardless, in sensitivity analyses in the Supplementary Information section entitled Phylogeny, we investigated this trade-off, and found that more than 1,000 species would have to be excluded from the data if higher-quality phylogenies were used (Supplementary Table 3). Further, inference is generally consistent across the datasets regardless of phylogeny quality (Supplementary Figs. 6 and 7).

After removing species not present in the Open Tree of Life topology, we further trimmed the data to include only higher-quality time series, removing the following: time series that contained zeros (which we considered extreme cases of extinctions or recolonizations) and time series with missing abundance values for a given year throughout the sampling duration (that is, we required consecutive abundance estimates). In all datasets except the two smallest ones—Atlantic reef fishes and Large carnivores—we further trimmed the datasets to keep only time series that had greater than or equal to the median number of abundance observations (that is, including the longest 50% of time series in each dataset). In some cases, this cutoff was not sufficient as the median number of observations in the time series equalled two. With only two abundance observations, trends are highly exposed to error purely driven by random fluctuations in abundance[19]. To partially address this issue, we imposed a further cutoff on these datasets, ensuring that each time series had at least five observations. These datasets are characterized in Supplementary Table 2. With our trimmed dataset, we derived a mean abundance in each year (in cases for which there were more than one observation per year) for each time series. In some datasets, there is a possibility that some species will have overlapping populations measured at different scales (for example, a national trend and a regional trend).

Modelling

We explored which models have been used in the literature to infer abundance change. We focused on studies that characterized the average change in abundances over time, rather than studies assessing how many species are declining or increasing, as this avoids discretizing a numeric value; that is, we avoid having to define what change is necessary to be classified as a decline. To evaluate the diversity of approaches used to model abundance change over time in multi-species and/or multi-location datasets, we conducted a literature search in a selection of high-profile ecology journals over the past 13 years (Supplementary Information). Our search identified 282 research papers, 28 of which described approaches to model abundance change across multi-species and/or multi-location datasets. A further 16 methods were not detected in the literature search but were known priori to the authorship team, resulting in 44 different studies and/or methods. Models of abundance change varied in complexity, each containing their own assumptions, with no clear 'standard' approach for deriving the rate of change in abundance. However, across the 44 studies and/or methods that we compiled (Supplementary Table 1), five general approaches were present, as follows.

1. Abundance average: The simplest models derive an average or total abundance across all species or sites in a given year, and then regress average abundance against time. This approach fails to recognize any of the hierarchical structures in the data.
2. Trend average: A slightly more complex model, which estimates abundance change per population by fitting a series of log-linear modes of abundance against year; averaging over the extracted slope coefficients. This approach fails to propagate uncertainty in average rates of change of each population, and ignores the implicit spatial and taxonomic structures in the data, inducing pseudoreplication.
3. Random intercept: Some studies partially address the aforementioned pseudoreplication (for example, certain sites or species having multiple estimates) with mixed models, regressing log-linear abundance against year across all populations, while specifying that populations belong to a site and/or species. However, often this mixed model structure extends only to random intercepts, which only acknowledge that mean abundance can differ between sites, species and locations, and assumes that the abundance trends will all remain the same. This is a particularly common approach among the indicators from population monitoring schemes that shape policy[46].
4. Random slope: In the scientific literature, it is common to use more complex models, with a similar structure to the random intercept model, but now also capturing the differences in abundance trends (not only mean abundances) across populations, sites and species with random slope terms.
5. Decomposition: This is the rarest of the approaches and deviates from the linear mixed model methods. Instead, the decomposition approach involves fitting generalized additive models through each time series to smooth abundance estimates and reduce noise. The smoothed time series is then decomposed into a time series of rates of change (or λ values), which are then averaged across species and biomes to derive estimates of the average change in abundance for each year across all of the time series.

The most common approaches were the random intercept and random slope models, used 19 and 23 times, respectively. The abundance average, trend average and decomposition approaches were rare, used just five, two and three times, respectively. Not all studies adopted just one approach, sometimes splitting their model into two steps (for example, using a random intercept model to estimate a given species trend across locations, which could then be aggregated across broader taxonomies with a random slope model). Further, all approaches regularly failed to account for temporal, spatial and phylogenetic structures (that is, closely related species are likely to have more similar trends

than distant species), with only 14 of the 44 approaches accounting for temporal autocorrelation. Studies that accounted for phylogenetic or spatial covariance were comparably rarer—included in just six and three studies, respectively. Four studies attempted to account for two sources of correlative non-independence in their models, by first deriving population trends while accounting for temporal autocorrelation of abundances in time series, and then using phylogenetic least squares to aggregate these trends. However, no study captured more than one of these covariances simultaneously (for example, spatio-temporal models). Further, no study attempted to account for all three sources of correlative non-independence.

Given the apparent rarity of the abundance average, trend average and decomposition approaches in the literature, we focus on understanding how the dominant approaches (that is, the random intercept and random slope models) compare to our newly developed correlated effect model. Full model equations are available in the Supplementary Information.

Model 1 (random intercept). In model 1, we fit a linear mixed-effect model between the natural logarithm of abundance and year, with five random intercepts: population (the unique time series), site (unique locations), region (broader spatial category to nest sites; flexibly determined on the basis of the spatial extent of the dataset), species (unique species) and genus (broader taxonomic category to nest species; measured as the parent node to the species tip). In the model, we do not specify any nesting of the population in the site and species random intercepts as the hierarchical structure of the data is poorly defined (for example, although populations always occur in a species and site, some species are nested in sites, and some sites are nested in species, creating a crossed random effect design). Model 1 assumes all populations, sites, regions, species and genera have the same trend in abundance.

Model 2 (random slope). In model 2, we develop a linear mixed-effect model, in which we regress the natural logarithm of abundance against year, including population, site, region, species and genus all as random slopes. This builds on the random intercept model by allowing abundance–year slope coefficients to vary for each category in each random slope term (for example, each species can have a different slope)—not simply differing intercepts as in model 1. Unlike in model 1, we centre the year and abundances of each population time series at zero (for example, subtracting each year by the mean year in each population, and subtracting the log of each abundance by the mean log abundance value in each population). This centring fixes the y and x intercepts at zero for each slope, and is a convenient solution to account for variance captured by the random intercepts without increasing the number of parameters. To all intents and purposes, assuming that the objective is to characterize the abundance–year coefficient, the random slope model is equivalent to a model with random intercepts and slopes.

Model 3 (correlated effect). Model 3 is structurally similar to model 2, but accounts for correlative non-independence. For temporal non-independence, we model the population level time series with a discrete first-order autoregressive temporal process, which assumes that sequential abundance observations in a time series will be more similar. To capture the spatial and phylogenetic correlative non-independence, we focus on non-independence across time-series trends (instead of abundance observations), assuming that trends in population abundances through time will be more similar in neighbouring sites and more closely related species. In models 1 and 2, we try to capture this non-independence with grouping categories (genus and region). However, in the correlated effect model, we more explicitly describe shared correlations between every species and site by specifying the covariance structures of our site and species random slopes. The site covariance matrix was derived by taking each site's coordinates and estimating the pairwise Haversine (spherical) distance between the sites (for example, how far is every site from every other site). This was then converted into a matrix, normalized between 0 and 1, with values close to 1 indicating neighbouring sites, whereas values approaching 0 indicate distant sites. The species covariance matrix was derived by converting the phylogeny into a variance–covariance matrix, in which phylogenetic branch lengths describe the evolutionary distance between species.

All models were developed using Bayesian Integrated Nested Laplace Approximation (INLA)[47] in R v.4.0.5 (ref. 48). We describe our model priors in the Supplementary Information section entitled Priors and validate our model assumptions in the Supplementary Information section entitled Assumptions (Supplementary Figs. 1–5). We also conduct additional sensitivity analyses exploring how phylogeny quality and how the addition of each correlative component (space, time or phylogeny) affect inference (see the Supplementary Information sections entitled Phylogeny and Component importance). We compiled the data using the following R packages: tidyverse[49], countrycode[50], janitor[51], here[52] and arrow[53]. Figures were produced using the following R packages: ggplot2[54], ggtree[55] and ape[45].

Outputs

Measuring non-independence. We assess whether correlative terms capture a meaningful proportion of variance in the data, by dividing the proportion of variance captured by the correlated slopes (for example, spatial covariance) by the combination of the variance captured by the correlated and independent slopes (for example, spatial covariance + site random slope + region random slope). This was carried out separately for the spatial and phylogenetic terms. As the spatial and phylogenetic components each contain three terms (an independent species or location slope, an independent genera or region slope and a correlated species or location slope), a proportional variance captured of 0.33 would indicate that the correlative slope captures an equal proportion of variance compared to the two independent slopes. A value greater than 0.33 indicates that correlative slopes account for more variation than independent random slopes. We measure temporal non-independence as the degree of correlation between sequential abundances (ρ).

Differing inference between the models. Using the mean and 50% credible intervals of the global trend (overall abundance–time coefficients), we display abundance projections for each model in each dataset. These projections are based on an arbitrary baseline abundance of 100, set at the first year of available data in each dataset, and this abundance would change according to the overall coefficients and credible intervals. For instance, with a 1% annual rate of change, an abundance in year zero of 100 would become 101 in year 1, and 164 in year 50. The purpose of these projections is to showcase varying abundance trajectories under different model complexities. We also report the fold change in the collective trend s.d. of the correlated effect model, relative to the random intercept and random slope models. This involved regressing fold change against category (for example, correlated effect versus random intercept) in a linear model. We report the mean fold change and 95% CIs. Model outputs are reported in Supplementary Tables 4 and 5.

Predictive performance. We assess the predictive performance of the different models by determining their ability to predict final observations in time series, and their ability to predict population trends of a given species in a given location. To test the predictive accuracy for the final observation in the time series, we removed the final observation from half of the time series in each dataset and predicted the missing values using each of the three models on the log scale. We report the percentage error (PE), a metric describing the median of the absolute percentage difference between predicted and observed values (for example, with a 5% error, an abundance on the log scale of 1 would

become 1.05). This is calculated by finding the absolute difference between the true value and prediction, divided by the true value, before being multiplied by 100 and converted to an absolute error.

To test the accuracy of the population trend prediction, we conducted leave-one-out cross-validation, removing one population time series (or trend) from each dataset, and estimating the missing trend using the random slope and correlated effect models. We solely removed population time series with a trend not overlapping zero at 95% credible intervals (that is, populations changing significantly), to test our ability to identify which populations are changing or not. We repeated this process 50 times per dataset and compare the predicted trends to trends from a simplified correlated effect model, which contains a population-level slope and accounts for temporal autocorrelation, but does not include the spatial and phylogenetic correlation terms or any of the hierarchical terms, which have no influence on the required population-level inference. We measured trend-predictive performance using the same approaches as above (PE). In the random slope model, the population trend coefficients were derived by adding the species, location, genus, region and overall coefficients together, meaning that missing population trends can still be informed by other hierarchical information. For the correlated effect model, the population trend is informed by the phylogenetic and spatial variance–covariance matrices, as well as all hierarchical information in the random slope model. Prediction accuracy for each dataset is reported in Supplementary Tables 6 and 7.

Phylogenetic and spatial distribution of abundance change. To plot abundance change across a phylogeny, we derived species-level rates of change in abundance from the taxonomic (species and genera) and phylogenetic random effects. We incorporate uncertainty in species-level trend prediction by estimating the CI threshold by which a species would be considered to have increased or decreased. For instance, a negative trend at an 80% CI threshold would be considered stronger evidence of decline than a negative trend at a 20% interval threshold. We derive four asymptotic CI thresholds (20%, 40%, 60% and 80%) using the uncertainty (s.d.) from the phylogenetic random effect and a series of *z*-scores (0.25, 0.52, 0.84 and 1.28).

To plot abundance change across space, we focus solely on one abundant and iconic species, the American robin *T. migratorius*, as site-level trend variability is high at the community level (that is, community trends across space are rarely significant). To produce abundance change predictions for the American robin across space, we expanded the BioTIME spatial Haversine distance matrix (describing distances between each time series) by supplementing it with a gridded extent covering North America. This new grid had a latitudinal range of 20 to 60 and 1° spacing (for example, 15, 16 and so on), and longitudinal range of −130 to −60 with 1° spacing. This new matrix allows us to estimate expected covariance (similarity) in abundance trends for any pair of 1° cells across North America. We then derived the average rate of change in abundance across all hierarchical and correlative random effects, and used population-level trend uncertainty to derive the selection of CI thresholds described above.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All of the data used in the study are publicly available and accessible from the following links: RivFishTIME (https://doi.org/10.1111/geb.13210), North American Breeding Birds (https://doi.org/10.5066/P97WAZE5), BioTIME (https://doi.org/10.1111/geb.12729), Living Planet (https://www.livingplanetindex.org/data_portal), CaPTrends (https://doi.org/10.1111/geb.13587), ReSurveyGermany (https://doi.org/10.25829/idiv.3514-0qsq70), UK Fish Counts (https://environment.data.gov.uk/dataset/ce2618db-d507-4671-bafe-840b930d2297), FishGlob (https://doi.org/10.31219/osf.io/2bcjw), TimeFISH (https://doi.org/10.1002/ecy.3966), Pilotto (https://zenodo.org/records/10638241). See comprehensive data descriptions in Supplementary Table 2 and data_compile.Rmd (https://zenodo.org/records/10638241). Source data are provided with this paper.

Code availability

All code is available at https://zenodo.org/records/10638241.

41. Jandt, U. et al. ReSurveyGermany: vegetation-plot time-series over the past hundred years in Germany. *Sci. Data* **9**, 631 (2022).
42. OpenTree. *Open Tree of Life Synthetic Tree version 13.4* https://doi.org/10.5281/zenodo.3937741 (2019).
43. Michonneau, F., Brown, J. W. & Winter, D. J. rotl: an R package to interact with the Open Tree of Life data. *Methods Ecol. Evol.* **7**, 1476–1481 (2016).
44. Grafen, A. & Hamilton, W. D. The phylogenetic regression. *Philos. Trans. R. Soc. B* **326**, 119–157 (1989).
45. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
46. *Breeding Bird Survey* (BTO, 2020, accessed May 1st 2023); https://doi.org/10.5066/P97WAZE5.
47. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B* **71**, 319–392 (2009).
48. R Core Team. *R: A Language and Environment for Statistical Computing*. http://www.R-project.org/ (R Foundation for Statistical Computing, 2020).
49. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
50. Arel-Bundock, V., Enevoldsen, N. & Yetman, C. countrycode: an R package to convert country names and country codes. *J. Open Source Softw.* **3**, 848 (2018).
51. Firke, S. *janitor: Simple Tools for Examining and Cleaning Dirty Data* https://sfirke.github.io/janitor/authors.html (2024).
52. Müller K. *here: A Simpler Way to Find Your Files* Version 1.0.1 https://here.r-lib.org/ (2020)
53. Richardson, N. et al. *arrow: Integration to 'Apache' 'Arrow' Version 15.0.0* https://github.com/apache/arrow/ (2024).
54. Wickham H. *ggplot2: Elegant Graphics for Data Analysis* https://ggplot2.tidyverse.org (Springer, 2016).
55. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

Acknowledgements We thank: the contributors to BioTIME; the Living Planet team, including R. Freeman and L. McRae; all scientists and volunteers that contributed to the North American Breeding Bird Survey; FishGlob; RivFishTIME; the UK environment agency; TimeFISH; ReSurveyGermany; as well as contributors to the European biodiversity data, including A. Verstraeten, J. Neirynck, G. Van Ryckegem, G. Van Hoey, B. Nikolov, V. Evtimova, R. Stanchev, P. Haase, I. Kronck, J. Meyer, J. Creveceour, M. Janssen, A. Thimonier, F. Pilotto, I. Kappel Schmidt, D. Oro, J. Poyry, K. Huttunen, T. Muotka, R. Paavola, L. Barboro, E. Camatti, M. Pansera, R. Alber, S. Vorhauser, A. Skula, G. Springe, B. J. Ens, M. E. Visser, T. Bongard, T. Jensen, K. A. E. Baekkelie, T. C. Jensen, M. Pardal, F. Martinho, T. Pipan, M. Aljancic, F. Gabrovsek, H. Kalivoda, L. Halada, S. David, U. Grandin, D. Monteith, S. Rennie, J. Adamson, R. Anderson, C. Andrews, J. Bater, N. Bayfield, K. Beaton, D. Beaumont, S. Benham, V. Bowmaker, C. Britt, R. Brooker, D. Brooks, J. Brunt, G. Common, R. Cooper, S. Corbett, N. Critchley, P. Dennis, J. Dick, B. Dodd, N. Dodd, N. Donovan, J. Easter, M. Flexen, A. Gardiner, D. Hamilton, P. Hargreaves, M. Hatton-Ellis, M. Howe, J. Kahl, M. Lane, S. Langan, D. Lloyd, B. McCarney, Y. McElarney, C. McKenna, S. McMillan, F. Milne, L. Milne, M. Morecroft, M. Murphy, A. Nelson, H. Nicholson, D. Pallett, D. Parry, I. Pearce, G. Pozsgai, R. Rose, S. Schafer, T. Scott, L. Sherrin, C. Shortall, R. Smith, P. Smith, R. Tait, C. Taylor, M. Taylor, M. Thurlow, C. Tilbury, A. Turner, K. Tyson, H. Watson, M. Whittaker, C. Wood, I. Winfield, M. I. Di Fonzo, B. Collen, A. L. M. Chauvenet, G. M. Mace, L. Comte, Q. J. Carvajal, P. A. Tedesco, X. Giam, U. Brose, T. Erős, A. F. Filipe, M. J. Fortin, K. Irving, C. Jacquet, S. Larsen, J. R. Sauer, D. K. Niven, J. E. Hines, D. J. Ziolkowski, K. L. Pardieck Jr, J. E. Fallon, W. A. Link, S. J. Thackeray, M. Dornelas, L. H. Antão, F. Moyes, A. E. Bates, A. E. Magurran, S. P. R. Greenstreet, M. Moriarty, S. D. Batten, S. Seibold, W. Weisser, M. Gossner, E. Pasalic, M. Lange, M. Turke, S. N. Gallenberger and M. Staab. This work was financially supported by the UK Research and Innovation–Natural Environment Research Council grant NE/T003502/1.

Author contributions Conceptualization: T.F.J., A.P.B., D.Z.C., R.P.F.; data curation: T.F.J., T.J.W., C.A.G.; analysis: T.F.J., P.C., R.P.F.; visualization: T.F.J.; writing: T.F.J., A.P.B., D.Z.C., R.P.F.; editing: T.F.J., A.P.B., D.Z.C., T.J.W., K.L.E., C.A.G., P.C., C.F.C., M.B., R.D.G., G.H.T., E.D., R.P.F.; mentoring: A.P.B., D.Z.C., R.P.F.; funding acquisition: A.P.B., R.P.F., D.Z.C., C.F.C., K.L.E., T.J.W., G.H.T.

Competing interests The authors declare no competing interests.

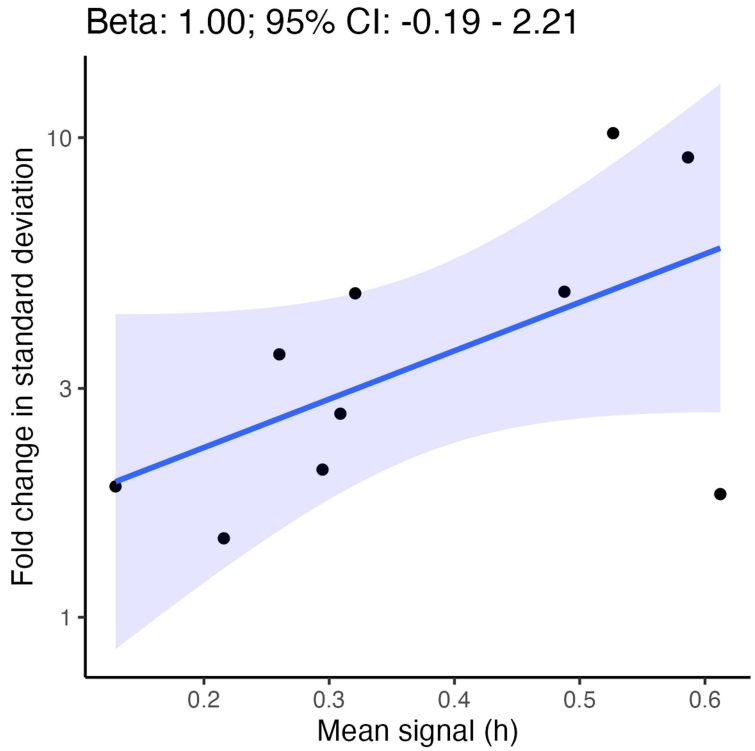
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07236-z.

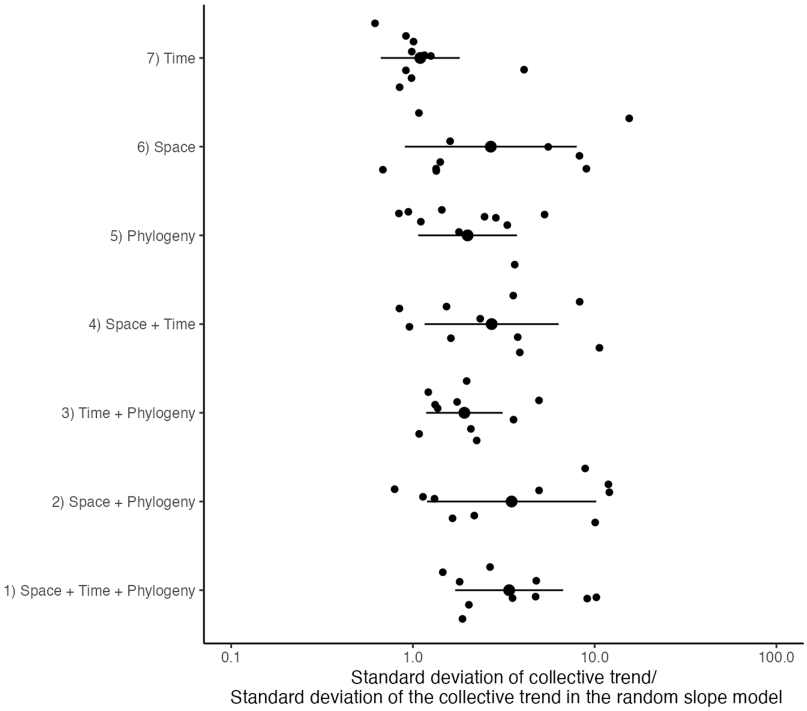
Correspondence and requests for materials should be addressed to T. F. Johnson.

Peer review information *Nature* thanks Henrique Pereira and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Impact of phylogenetic and spatial signal on inference. Fold change in the standard deviation of the abundance-time coefficient between the correlated effect and random slopes model, plotted against the mean signal. Mean signal is calculated by finding the mean of the phylogenetic signal (variance captured by the phylogeny divided by the sum of the phylogenetic and non-phylogenetic variance) and spatial signal (as in the phylogeny). n = 10 with each point representing a dataset. Shading represents 95% confidence intervals from a linear model.



Extended Data Fig. 2 | Contribution of spatial, temporal and phylogenetic components to collective trend uncertainty. Relative fold change in the standard deviation of the collective trend as additional components (space, time or phylogeny) are added to the random slope model. Models are compared to the standard deviation of the collective trend in the random slope model. In each model comparison (y-axis), n = 10 with each point representing a dataset. The larger point and error bar represents the mean change and associated standard deviation around this mean.

nature portfolio

Corresponding author(s): Thomas Frederick Johnson

Last updated by author(s): Feb 7, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

We have a fully reproducible and annotated data collection pipeline in the form of an RMarkdown document, named 'data_compile.Rmd' within our code repository (https://zenodo.org/records/10638241). Data was compiled from 10 sources (seed data availability below) and processed within R V4.0.5 using the following packages: tidyverse 2.0.0, here 1.0.1, janitor 2.2.0, countrycode 1.5.0 and arrow 12.0.1.

Arel-Bundock et al. (2018) countrycode: An R package to convert country names and country codes. Journal of Open Source Software https://doi.org/10.21105/joss.00848. Version 1.5.0

Firke S. (2023) janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.2.0, https://CRAN.R-project.org/package=janitor

Müller K. (2020) here: A Simpler Way to Find Your Files. R package version 1.0.1, https://CRAN.Rproject.org/package=here

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Version 4.0.5

Richardson, N. et al. (2023) arrow: Integration to 'Apache' 'Arrow'. R package version 11.0.0.3, https://CRAN.R-project.org/package=arrow

Wickham H. et al. (2019) Welcome to the tidyverse. Journal of Open Source Software https://doi.org/10.21105/joss.01686. Version 2.0.0

Data analysis

All of our code is openly available and reproducible from https://zenodo.org/records/10638241. Here, we provide a summary of our analysis, more information is available in the main text, methods, supplementary material and code.

The objective of our study was to consider how inference made in commonly applied biodiversity models could be impacted by failing to account for spatial, temporal and phylogenetic dependencies. Prior to designing our models, we first explored what models have been used in the literature to infer abundance change. In this work, we focussed on studies trying to characterise the average change in abundances over time, rather than studies attempting to assess how many species are declining or increasing, as this avoids discretizing a numeric value i.e. by assessing the average change, we avoid having to define what change is necessary to be classified as a 'decline'. To evaluate the diversity of approaches used to model abundance change over time in multi-species and/or multi-location datasets, we conducted a literature search within a selection of high profile ecology journals over the last 13 years. Our search identified 282 research papers, 28 of which described approaches to model abundance change across multi-species/location datasets. A further 16 methods were not detected within the literature search but were known priori to the authorship team, resulting in 44 different studies/methods. Models of abundance change varied in complexity, each containing their own assumptions, with no clear 'standard' approach for deriving the rate of change in abundance. However, across the 44 studies/methods we compiled (Table 1), five general approaches were present (Table 1):

Abundance average - The simplest models derive an average or total abundance across all species or sites in a given year, and then regress average abundance against time. This approach fails to recognise any of the hierarchical structure in the data.

Trend average - A slightly more complex model, which estimates abundance change per population by fitting a series of log-linear modes of abundance against year; averaging over the extracted slope coefficients. This approach fails to propagate uncertainty within average rates of change of each population, and ignores the implicit spatial and taxonomic structure within the data, inducing pseudoreplication.

Random intercept - Some studies partially address the aforementioned pseudoreplication (e.g. certain sites or species having multiple estimates) with mixed models, regressing log-linear abundance against year across all populations, whilst specifying that populations belong to a site and/or species. However, often this mixed model structure only extends to random intercepts , which only acknowledge that mean abundance can differ between sites, species and location, but assumes that the abundance trends will all remain the same. This is a particularly common approach amongst the indicators from population monitoring schemes which shape policy.

Random slope - In the scientific literature, it is common to use more complex models, with a similar structure to the Random intercept model, but now capturing the differences in abundance trends across populations, sites and species with random slope terms.

Decomposition - This is the rarest of the approaches and deviates from the linear mixed model approaches. Instead, the decomposition approach involves fitting generalised additive models (GAMs) through each time series to smooth abundance estimates and reduce noise. The smoothed time series is then decomposed into a timeseries of rates of change (or lambdas), which are then averaged across species and biomes to derive estimates of the average change in abundance for each year across all the time-series.

The most common approaches were the random intercept and random slope models, used 19 and 23 times, respectively. The abundance average, trend average and decomposition approaches were rare, used just 5, 2, and 3 times, respectively. Not all studies adopted just one approach, sometimes splitting their model into two steps e.g. using a random intercept model to estimate a given species trend across locations, which could then be aggregated across broader taxonomies with a random slope model. Further, all approaches regularly failed to recognise that abundance patterns are shaped by implicit temporal, spatial and phylogenetic signals (i.e. closely related species are likely to have more similar trends than distant species), with only 14 of the 44 approaches accounting for temporal autocorrelation. Phylogenetic and spatial covariance were comparably rarer - included in just 6 and 3 studies respectively. Four studies attempted to account for two sources of correlative non-independence within their models, by first deriving population trends whilst accounting for temporal autocorrelation of abundances within time series, and then using phylogenetic least squares to aggregate these trends. However, no study captured more than one of these covariances simultaneously (e.g. spatio-temporal models for instance). Further, no study attempted to account for all three sources of correlative non-independence at the same time.

Given the apparent rarity of the abundance average, trend average and decomposition approaches within the literature, we focus on understanding how the dominant approaches (i.e. the random intercept and random slope models) compare to our newly developed correlated effect model. Full model equations are available in 'Supplementary material - Models'.

Model 1. Random intercept

In model 1, we fit a linear mixed effect model between the natural logarithm of abundance and year, with five random intercepts: population (the unique time series), site (unique locations), region (broader spatial category to nest sites; measured as 10-degree grid cell the site occurs in), species (unique species), and genus (broader taxonomic category to nest species; measured as the parent node to the species tip). Within the model, we do not specify any nesting of the population within the site and species random intercepts as the hierarchical structure of the data is poorly defined e.g., whilst populations always occur within a species and site, some species are nested in sites, and some sites are nested in species, creating a crossed random effect design. Model 1 assumes all populations, sites, regions, species, and genera have the same trend in abundance.

Model 2. Random slope

In model 2, we develop a linear mixed effect model, where we regress the natural logarithm of abundance against year, including population, site, region, species, and genus all as random slopes. This builds on the random intercept model by allowing abundance-year slope coefficients to vary for each category in each random slope term (e.g., each species can have a different slope) - not simply differing intercepts as in model 1. Unlike model 1, we centre the year and abundances of each population time series at zero e.g. subtracting each year by the mean year in each population, and subtracting the log of each abundance by the mean log abundance value in each population. This centering fixes the y and x intercepts at zero for each slope, and is a convenient solution to acknowledge variance captured by the random intercepts without increasing the number of parameters. In all intents and purposes, the random slope model is equivalent to a model with random intercepts and slopes.

Model 3. Correlated effect

Model 3 is structurally similar to model 2, but accounts for correlative non-independence structures. For temporal non-independence, we model the population level time series with a discrete autoregressive-1 (ar-1) temporal process, which assumes neighbouring abundance observations within a time series will be more similar. To capture the spatial and phylogenetic correlative non-independence, we focus on non-independence across time series trends (instead of abundance observations), assuming trends in population abundances through time will be more similar in neighbouring sites and more closely related species. In model 1 and 2, we try to capture this non-independence with grouping categories (genus and region). However, in the correlated effect model, we more explicitly describe shared correlations between every species and site by specifying the covariance structure of our site and species random slopes. The site covariance matrix was derived by taking each site's coordinates and estimating the pairwise Haversine (spherical) distance between the sites e.g. how far is every site from every other site. This was then converted into a matrix, normalised between 0 and 1, with values close to 1 indicating neighbouring sites, whilst values approaching 0 indicate distant sites. The species covariance matrix was derived by converting the phylogeny into a variance-covariance matrix, where phylogenetic branch lengths describe the evolutionary distance between species.

All models were run in INLA using R V4.0.5. We describe our model priors in 'Supplementary material - Priors' and validate our model assumptions in 'Supplementary material - Assumptions'. We also conduct additional sensitivity analyses exploring how phylogeny quality and how the addition of each correlative component (space, time or phylogeny) impacts inference - see 'Supplementary material - Phylogeny' and 'Supplementary material - Component contribution'.

Outputs

Measuring non-independence
In our Correlated effect model we measure the presence of total non-independence as the proportion of variance captured by the combination of independent and correlative terms (i.e. random effects) for each component (e.g., temporal components), divided by the sum of the variance for all terms. Next, we assess if correlative terms are the larger contributor to this total non-independence, by dividing the proportion of variance captured by the correlated slopes, by the combination of the variance captured by the correlated and independent slopes. This was done separately for the spatial and phylogenetic terms. As the spatial and phylogenetic components each contain three terms (an independence species/location slope, an independent genera/region slope, and a correlated species/location slope), a proportional variance captured of 0.33 would indicate that the correlative slope captures an equal proportion of variance compared to the two independent slopes. A value greater than 0.33 indicates that correlative slopes account for more variation than independent random slopes. We measure temporal non-independence as the degree of correlation between neighbouring abundances (rho).

Differing inference between the models
Using the mean and 50% credible intervals of the global trend (overall abundance-time coefficients), we display abundance projections for each model in each dataset. These projections are based on an arbitrary baseline abundance of 100, set at the first year of available data in each dataset, and this abundance would change according to the overall coefficients and credible intervals. For instance, with a 1% annual rate of change, an abundance in year zero of 100, would become 101 in year 1, and 164 in year 50. The purpose of these projections is to showcase varying abundance trajectories under different model complexities.

Next, we note the number of the datasets where inference reverses (e.g., the global trend reverses direction from positive to negative, or remains consistent), and where uncertainty increases (the variance around the global trend is greater or smaller), comparing the random intercept and random slope models to the correlated effect model. To support these comparisons, we also report the fold change in the collective trend standard deviation of the correlated effect model, relative to the random intercept and random slope models.

Predictive performance
We assess the predictive performance of the different models by determining their ability to predict final observations in time series', and their ability to predict population trends of a given species in a given location. To test the predictive accuracy for the final observation in the time series, we removed the final observation from half of the time-series in each dataset and predicted the missing values using each of the three models on the log scale. We report the percentage error (PE), a metric describing the median of the absolute percentage difference between predicted and observed values e.g. with a 5% error, an abundance on the log-scale of 1 would become 1.05. To test the accuracy of the population trend prediction, we conducted leave-one-out cross validation, removing one population time series (or trend) from each dataset, and estimating the missing trend using the random slope and correlated effect models. We repeated this process 50 times per dataset and compare the predicted trends to trends from a simplified correlated effect model, which contains a population level slope and accounts for temporal autocorrelation, but does not include the spatial and phylogenetic correlation terms or any of the hierarchical terms, which have no influence on the required population level inference. We measured trend predictive performance using the same approaches as above (PE). We estimate this error statistic across the full sample of 50 populations per dataset, and across the sub-sample of the rarest 20% of populations i.e. populations in the least well-studied species and location. In the random slope model, the population trend coefficients were derived by adding the species, location, genus, region, and overall coefficients together, meaning missing population values can still be informed by other hierarchical information. For the correlated effect model, the population trend is informed by the phylogenetic and spatial variance-covariance matrices, as well as all hierarchical information in the random slope model.

Phylogenetic and spatial distribution of abundance change
To plot abundance change across a phylogeny, we derived species level rates of change in abundance from the taxonomic (species and genera) and phylogenetic random effects. We incorporate uncertainty in species-level trend prediction by estimating the confidence interval threshold by which a species would be considered to have increased or decreased. For instance, a negative trend at an 80% confidence interval threshold would be considered stronger evidence of decline than a negative trend at at 20% interval threshold. We derive four asymptotic confidence interval thresholds (20%, 40%, 60%, 80%) using the uncertainty (standard deviation) from the phylogenetic random effect and a series of z-scores (0.25, 0.52, 0.84, 1.28).

To plot abundance change across space, we focus solely on one abundant and iconic species, the American Robin Turdus migratorius, as site-level trend variability is high at the community level i.e. community trends across space are rarely significant. To produce abundance change predictions for the American Robin across space, we expanded the BioTIME spatial Haversine distance matrix (describing distances between each time series) by supplementing it with a gridded extent covering North America. This new grid had a latitudinal range of 20 to 60 and 1 degree spacing (e.g. 15, 16, etc.), and longitudinal range of -130 to -60 with 1 degree spacing. This new matrix allows us to estimate expected

covariance (similarity) in abundance trends for any pair of 1 degree cells across North America. We then derived the average rate of change in abundance across all hierarchical and correlative random effects, and used population-level trend uncertainty to derive the selection of confidence interval thresholds described above.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All of the data used in the study are publicly available and accessible from the following links: RivFishTIME (https://doi.org/10.1111/geb.13210), North American Breeding Birds (https://doi.org/10.5066/P97WAZE5), BioTIME (https://doi.org/10.1111/geb.12729), Living Planet (https://www.livingplanetindex.org/data_portal), CaPTrends (https://doi.org/10.1111/geb.13587), ReSurvey Germany (https://doi.org/10.25829/idiv.3514-0qsq70), UK Fish Counts (https://environment.data.gov.uk/dataset/ce2618db-d507-4671-bafe-840b930d2297), FishGlob (https://doi.org/10.31219/osf.io/2bcjw), TimeFISH (https://doi.org/10.1002/ecy.3966), Pilotto (https://zenodo.org/records/10638241). See below for a summary of the data sources:

1) RivFishTIME

RivFishTime is a global database of freshwater fish time-series to study global change ecology in riverine systems, fully described in [Comte et al. 2020](https://doi.org/10.1111/geb.13210), with the data hosted by iDiv [here](https://idata.idiv.de/ddm/Data/ShowData/1873?version=12). Please see the full description and meta data of this database for more information.

The full zipped database can be downloaded directly and unzipped (before removing the zipped version). The two files required are the main survey dataset '1873_2_RivFishTIME_SurveyTable.csv' (this contains the record-level data, i.e. abundance of a given species at a given site in a given year), and the dataset describing each individual time series '1873_2_RivFishTIME_TimeseriesTable.csv' (this contains information on the specific location of each time series).

2) North American Breeding Bird Survey

The 2022 release of the North American Breeding Bird Survey dataset (1966-2021) is available from [Ziolkowski Jr. et al. (2022)](https://doi.org/10.5066/P97WAZE5). From the dataset description there: This dataset contains avian point count data for more than 700 North American bird taxa (species, races, and unidentified species groupings), collected annually during the breeding season along thousands of randomly established roadside survey routes in the United States and Canada. Routes are roughly 24.5 miles (39.2 km) long with counting locations placed at approximately half-mile (800-m) intervals, for a total of 50 stops. At each stop, a citizen scientist highly skilled in avian identification conducts a 3-minute point count, recording all birds seen within a quarter-mile (400-m) radius and all birds heard. Surveys begin 30 minutes before local sunrise and take approximately 5 hours to complete. Routes are sampled once per year, with the total number of routes sampled per year growing over time; just over 500 routes were sampled in 1966, while in recent decades approximately 3000 routes have been sampled annually. No data are provided for 2020. BBS field activities were cancelled in 2020 because of the coronavirus disease (COVID-19) global pandemic and observers were directed to not sample routes. Route location information includes country, state, and BCR, as well as geographic coordinates of route start point, and an indicator of run data quality. We require the 'States' and 'Routes' zipped datasets, and the species list (provided as a text file). The states data are provided as a zipped file for each state, but each of these is just a single csv file so can be read directly with `read_csv` To read them all into one big dataframe (specifying which columns to read, and stating their required data types). Processing involves adding location details and species names to this large dataset, creating a unique 'site' ID (from country, state, and route information), and mutating, renaming, or adding the other required variables:

3) BioTIME

BioTIME [(Dornelas et al. 2018)](https://doi.org/10.1111/geb.12729) is a comprehensive collection of assemblage time-series in which the abundances of the species that comprise ecological communities have been monitored over a number of years. BioTIME data span the globe and encompass land and seas; they also include freshwater systems. The current version of BioTIME contains over 12 million records, features almost 50 thousand species, covers over 600 thousand distinct geographic locations and is representative of over 20 biomes, occurring over 6 different climatic zones. This dataset requires registration prior to download. We downloaded the June 2021 version (the latest available) of the raw data in CSV format, together with the meta data and citations files, from the link provided following registration at https://biotime.st-andrews.ac.uk/download.php.
Processing the BioTIME data requires selecting relevant columns, and joining to the metadata in order to filter out studies that recorded only presence/absence, or only biomass (no abundance data). A 'site' variable is created by pasting the study ID, latitude, and longitude. No country information is provided.

4) Living Planet Index

The Living Planet Index (LPI; [LPI 2022](www.livingplanetindex.org/)) is a measure of the state of the world's biological diversity based on population trends of vertebrate species from terrestrial, freshwater and marine habitats. The LPI is based on trends of thousands of population time series collected from monitored sites around the world. Accessing the dataset requires registration prior to download from https://www.livingplanetindex.org/data_portal. We downloaded the latest available (2022) zipped version of the database into our raw data folder.

The subfolder 'LivingPlanetIndex_2022_PublicData' includes the LPI data agreement (data_agreement_2022.pdf) and metadata (LPD_metadata.pdf) as well as the public data as a csv file (LPD2022_public.csv). Read in this csv - NB 'NULL' is used to indicate missing abundance values. Country names are supplied, to convert

these to 3 character ISO codes we use the `countrycode` package [(Arel-Bundock et al. 2018)](https://doi.org/10.21105/joss.00848). There are two countries in the LPI data that do not match due to character encoding issues, plus 'International Waters', so we set up custom matches for these.

5) CaPTrends

CaPTrends ([Johnson et al. 2022](https://doi.org/10.1111/geb.13587)) is a database of 1,122 population trends from around the world, describing changes in abundance over time in large mammal species (n =50) from four families (Canidae, Felidae, Hyaenidae and Ursidae) in the order Carnivora. Trends represent 621 unique locations across the globe (latitude: −51.0 to 80.0; longitude: −166.0 to 166.0), from 1726 to 2017. The dataset itself is hosted on Zenodo here: https://zenodo.org/record/6949487. First, download the zipped dataset and unzip. Examining the meta data and descriptions, the two files we require are 'captrends.csv' (details of each individual study) and 'abundance.csv' (the actual abundance values). Some studies cover multiple countries - we assign the code to the first country mentioned. A 'site' variable is created by pasting `data_table_id`, `citation_key`, and `locality_name`. Latitude and longitude values are not available so these are set to NA. The abundance unit is set to the value in `population_metric` if available, otherwise to the `field_method` value.

6) ReSurvey Germany: vegetation-plot resurvey data from Germany

ReSurvey Germany: vegetation-plot resurvey data from Germany ([Jandt et al. 2022](https://doi.org/10.25829/idiv.3514-0qsq70)) is a compilation of harmonised vegetation-plot resurvey data from Germany covering almost 100 years. The data allow calculating temporal biodiversity change at the community scale. They also enable tracking changes in the incidence and distribution of individual species across Germany. Cover records are available for 1,794 vascular plant species in 7,738 (semi-)permanent vegetation plots from Germany, resurveyed from 2 to 54 times, in total resulting in 23,641 vegetation records and 458,311 species cover records, comprising the years from 1927 to 2020 and 97 EUNIS habitat types. The data is available to download from: https://doi.org/10.25829/idiv.3514-0qsq70. The main files required are the 'header' data (project / site level data) in 'Header_ReSurveyGermany.csv', and the main dataset of species-level abundances in 'ReSurveyGermany.csv'.

7) Environment Agency NFPD Fish Counts

The Environment Agency undertakes fisheries monitoring work on rivers, lakes and transitional and coastal waters (TraC) throughout England. The freshwater fish survey dataset (or _National Fish Populations Database_, NFPD, [Environment Agency 2020](https://environment.data.gov.uk/dataset/ce2618db-d507-4671-bafe-840b930d2297)) contains site and survey information, as well as the numbers and species of fish caught, for all the freshwater fish surveys carried out across England (with a small number from Wales and Scotland) from 1975 onwards. We accessed the Freshwater Fish Count dataset from https://environment.data.gov.uk/ecology/explorer/downloads/. Meta-data is available in a pdf document [here](https://environment.data.gov.uk/portalstg/sharing/rest/content/items/1150f6994d294d78b422b97848c3a286/data). The geographic coordinates in this database are provided as eastings and northings in the UK national projection ([EPSG 27700](http://www.opengis.net/def/crs/EPSG/0/27700)). In addition, there are some location errors placing surveys offshore or with incomplete coordinates (e.g. easting and northing both = 1). To address these, we use the `sf` library ([Pebesma 2018](https://doi.org/10.32614/RJ-2018-009)) to convert the database into spatial format, reproject to WGS84 latitude and longitude coordinates, and we combine it with a UK coastline map to exclude surveys that occur at sea. First, load `sf` and make the NFPD dataset spatial. We obtained the digital vector boundaries for Countries in the United Kingdom as at December 2022 at full resolution, clipped to the coastline (Mean High Water mark), from the Office for National Statistics, available https://geoportal.statistics.gov.uk/datasets/ons::countries-december-2022-uk-bfc/. We downloaded the shapefile, 'Countries_December_2022_UK_BFC_3731595901038458592.zip', directly downloaded and unzipped.

Now we can process this dataset. For the 'site' variable, there are two levels of site identification - the local survey site (identified by `site_id`, typically on the order of 100m river length) and a larger scale 'parent' site (`site_parent_id`), which may constitute several surveys within the same local area (typically all within a few km of each other). We retain both of these in a composite 'site' variable, meaning that the data are at the level of the site but that aggregation to parent site remains possible if required. Abundance is measured as the total number of fish sampled. This is done over one or more survey runs at each site in each time period. For around half the surveys, estimates of total population density are available by using the Carle & Strub equation ([Carle & Strub 1978](https://doi.org/10.2307/2530381)) over a three run catch depletion survey. However, this method can only be used on multiple run surveys, which would result in discarding over half available surveys, and so we take as our abundance measure the counts from the first run of the multiple run surveys, or the only run from single run surveys (as is done for Water Framework Directive classification; Philip Rudd, Fisheries Technical Specialist, Environment Agency, pers. comm. April 2023). Counts are then divided by the survey area (ength of river fished multiplied by the average width) to give abundance as individuals per 100 m^2^. For some surveys, exact counts are not given - these are excluded. Zero catches are recorded in a separate variable; these are converted to 0s in our abundance variable.

8) FishGlob Global Bottom Trawl Survey Database

FishGlob_data is a global database of bottom trawl survey data for marine fish, described in [Maureaud et al. (2023)](https://doi.org/10.31219/osf.io/2bcjw). The database contains a cleaned collation of 26 publicly available bottom-trawl surveys conducted in national waters of 18 countries that are standardised and pre-processed, covering a total of 2,162 sampled fish taxa and 232,800 hauls collected from 1963 to 2020. The database is available from Zenodo at https://zenodo.org/record/7527447#.ZDhrFuzMIqt. The full clean standardised dataset is found in the 'ouputs' subfolder as a .RData binary file, 'FishGlob_std_public_clean.RData'. This contains site level estimates of abundance per km2 from multiple bottom trawl surveys across multiple species. This data includes two objects, the main data as `data`, and meta-data in `readme`. Country is included in the dataset, but as country name. To convert to country code, create a dataframe of distinct countries, and use the `countrycode` ([Arel-Bundock et al. 2018](https://doi.org/10.21105/joss.00848)) package to obtain relevant ISO3 codes (all surveys listed as 'multi-countries' are in Europe, so we use EUROPE as a code for these).

Now process the data. As these abundances are derived through trawling, there is no discrete spatial repetition in sampling i.e. the exact same site is not sampled every year. Instead, the trawlers collecting data can deviate slightly from the exact site, but do often stay within the same general region/area. To handle these discrepancies in sampling location, we upscale the sampling locations (latitude and longitude) to a 1-degree resolution and take the average abundance estimates from the same sampling scheme,for each species in each year within this 1-degree gird cell. Simply put, we upscale the spatial resolution to allow temporal comparisons in trawling data. Note: A 1-degree resolution was decided on as it allowed us to develop these temporal comparisons whilst maintaining spatial structure

9) TimeFISH

The TimeFISH database ([Quimbayo et al. 2022](https://doi.org/10.1002/ecy.3966)) provides the first public time-series dataset on reef fish assemblages in the southwestern Atlantic (SWA), comprising 15years of data (2007–2022) based on standardized Underwater Visual Censuses (UVCs) in nine locations along the southern Brazilian coast (25–29°S). All fish individuals in the water column (up to 2m above the substratum) and at the bottom were targeted. In total, 202,965 individuals belonging to 163 reef fish species and 53 families were recorded across 1857 UVCs. Data are available to use with no restrictions, and can be downloaded from Zenodo: https://zenodo.org/record/7317084#.ZFJRK-zMIqs

Process the data. First, aggregate census data to species per transect (as we do not need individual size class abundances). Then fomat species names, and join to location data. Note - two transect IDs appear twice in the location data. As we are not interested in dates beyond years, we can disregard this in joining below by setting `multiple = "first"`. Then add dataset_id, create new site variable, add country code and unit, rename variables as needed and select the final set:

10) Pilotto

In 2020, Pilotto published a study on biodiversity trends in Europe, largely relaying on Europe's network of 'long-term ecological research sites'. Unlike the other datasets compiled in this Rmarkdown, Pilotto does not provide the full dataset, instead only sharing links to the raw data. To make use of the Pilotto data, we compiled this raw data into a series of .csv files (https://zenodo.org/records/10638241). Specifically, this involved working through the links available in Pilotto's data, downloading the raw data from these links into the directory 'data/raw_data/Pilotto/unique_id'. With each downloaded dataset, we then extracted information into three .csv files: compile_ts.csv - this file contains the abundance time-series, reported taxa, year, and name of the site; compile_sp.csv - this file contains the coordinates linked to sites contained in compile_ts.csv; and compile_tx.csv - this file contains species names, and is neccasary as reported taxa in 'compile_ts.csv' sometimes are described with codes instead of their actual name e.g. 'species_xyz123'. 'compile_tx.csv' can be used to convert these codes into actual species names. In some cases, we decided the information used by Pilotto was not suitable for our study (e.g. sometimes data contained presence/absence instead of abundance values). We have carefully annotated the data we extracted from Pilotto within the file 'Pilotto/master.csv'.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We assess how temporal trends in abundance change after accounting for spatial, temporal and phylogenetic dependencies. We test this across 10 high-profile datasets, reporting the change in abundance ~ year coefficient, and credible intervals. We also explore predictive accuracy after incorporating spatial, temporal and phylogenetic dependencies
Research sample	We use 10 high-profile abundance datasets, representing more than 30,000 populations, ~2,900 species and ~5,850 unique locations. For each dataset, we extracted the population abundance estimates, the accompanying time-stamps, the species scientific names, the name of the site (location) where the population was sampled, and any site coordinates. For datasets to be included they had to be open access, and contain multiple abundance time series for a selection of species and locations. Whilst these datasets are vital within biodiversity science, many of the datasets are prone to biases e.g. lacking tropical representation, and contain few plant and invertebrate species. The datasets have been compiled from a variety of methods, realms and systems, covering a vast array of spatial, taxonomic and temporal scales. Further, there is likely some overlap in data between datasets - i.e. population time-series may occur in more than one dataset. We take no action to correct or acknowledge these biases and features, as our analysis is designed to show how model choice can have a substantial influence on inference in a variety of datasets, rather than to derive new trend estimates for each dataset or derive a consensus trend across datasets.
Sampling strategy	We selected datasets with open-use policies to enable the work to be reproducible. We conducted no power analyses to determine sample sizes as our work included all available data.
Data collection	See comprehensive instructions on data collection above.
Timing and spatial scale	The datasets vary in temporal, spatial and taxonomic scale. Most records tend to occur within later decades, the global north, and vertebrate populations. Please see a summary for each dataset below Population abundance time series from the BioTIME dataset - representing all core taxa and realms. Covering 12,065 abundance time

series, derived from 243,993 abundance observations. These time series represent 438 unique sites and 1,233 species.
Temporal extent: 1933-2018
Latitude extent: -77.6 - 67.8
Longitude extent: -179.8 - 179.2

Global population abundance time series from the Living Planet database. Covering 5,255 abundance time series, derived from 166,827 abundance observations. These time series represent 1,159 unique sites and 1,264 species.
Temporal extent: 1950 - 2020
Latitude extent: -77.8 - 78.9
Longitude extent: -180 - 180

Population abundance time series from the North American breeding bird survey. Covering 8,718 abundance time series, derived from 164,317 abundance observations. These time series represent 584 unique sites and 361 species.
Temporal extent: 1966 - 2019
Latitude extent: 25.9 - 67.0
Longitude extent: -165.3 - -55.4

Population abundance time series from the FishGlob database, describing abundances from the bottom-trawl survey for marine fishes. Covering 2,286 abundance time series, derived from 67,908 abundance observations. These time series represent 229 unique sites and 152 species.
Temporal extent: 1977 - 2020
Latitude extent: 26 - 62
Longitude extent: -178 - 21

Population abundance time series from the RivFishTIME database. Covering 2,386 abundance time series, derived from 40,834 abundance observations. These time series represent 197 unique sites and 191 species.
Temporal extent: 1975 - 2019
Latitude extent: -28.3 - 67.9
Longitude extent: -122.4 - 153.4

Population abundance time series from the UK Environment Agency Fish population database, describing fish populations in rivers, lakes and transitional/coastal waters. Covering 361 abundance time series, derived from 3,016 abundance observations. These time series represent 181 unique sites and 16 species.
Temporal extent: 1984 - 2019
Latitude extent: 50.4 - 55.4
Longitude extent: -3.9 - 0.5

Population abundance time series from the TimeFISH database, describing abundances of reef assemblages in the South-western Atlantic. Covering 86 abundance time series, derived from 262 abundance observations. These time series represent 12 unique sites and 52 species.
Temporal extent: 2008 - 2022
Latitude extent: -27.7 - -27.1
Longitude extent: -48.5 - -48.3

Population abundance time series from the ReSurveyGermany database, describing relative cover in vegetation plots. Covering 356 abundance time series, derived from 4,954 abundance observations. These time series represent 7 unique sites and 93 species.
Temporal extent: 1965 - 2018
Latitude extent: 48.3 - 53.6
Longitude extent: 7.4 - 13.9

Population abundance time series from the Pilotto et al., (2020) study ''Meta-analysis of multidecadal biodiversity trends in Europe' dataset - representing diverse taxa across the terrestrial, freshwater and marine realms. Covering 2,386 abundance time series, derived from 40,834 abundance observations. These time series represent 197 unique sites and 191 species. Note: The compiled form of this dataset was not openly available (unlike all others), instead the dataset only provided the references to the primary sources. We extracted relative abundance/density estimates from across the 51 primary sources referenced within the database. We excluded a further 31 datasets contained within this database, as the data lacked clear metadata, or data were not resolved to the species level, or data represented species presence/absences instead of abundances.
Temporal extent: 1975 - 2019
Latitude extent: 40.1 - 67.8
Longitude extent: -8.9 - 29.6

Population abundance time series from the CaPTrends database of large carnivore population trends and time series. Covering 279 abundance time series, derived from 2,670 abundance observations. These time series represent 165 unique sites and 26 species.
Temporal extent: 1880 - 2019
Latitude extent: -40.0 - 71.6
Longitude extent: -158.0 - 99.2

Data exclusions

For each dataset we extract synthetic trees from the open tree of life 45,46 and estimate missing branch lengths using Grafen's approach 47 from the compute.brlen function in the R package ape 48. The Open Tree of Life identified a phylogeny for 80% of species (N = 23,871); all other species were removed from the analysis. For studies with the overall aim of assessing biodiversity change, removing species could be problematic, as the collective trend would not be representative of all species. However, in our case, where the aim is to assess how collective trend inference changes under a variety of modelling approaches, trimming the data to species with an accompanying phylogeny has no impact on our conclusions.

After removing species not present in the Open Tree of Life topology, we further trimmed the data to only include higher-quality time series, removing the following: time series that contained zeros (which we considered extreme cases of extinctions or recolonisations) and time series with missing abundance values for a given year throughout the sampling duration (i.e., we required consecutive abundance estimates.) In all datasets except the two smallest - G) Atlantic reef fishes & J) Large carnivores) - we further trimmed the datasets to only keep time series which had greater than or equal to the median number of abundance observations i.e., including the longest 50% of time series in each dataset. In some cases, this cut-off was not sufficient as the median number of observations in the time series equaled two. With only two abundance observations, trends are highly exposed to error purely driven by random fluctuations in abundance 10. To partially address this issue, we imposed a further cut-off on these datasets, ensuring each time-series had at least 5 observations. These datasets are characterised in Table S1. With our trimmed dataset, we derived a mean abundance in each year (in cases where there were more than one observation per year) for each time series.

Reproducibility: All code and code are openly avalable and reproducible

Randomization: Data was not randomized. Our study showcases the need to conduct statistical controls for non-independence, which has been been neglected in previous work

Blinding: NA

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging