Article

Clustered versus catastrophic global vertebrate declines

Brian Leung[1,2 ✉], Anna L. Hargreaves[1], Dan A. Greenberg[3], Brian McGill[4,5], Maria Dornelas[6] & Robin Freeman[7]

Recent analyses have reported catastrophic global declines in vertebrate populations[1,2]. However, the distillation of many trends into a global mean index obscures the variation that can inform conservation measures and can be sensitive to analytical decisions. For example, previous analyses have estimated a mean vertebrate decline of more than 50% since 1970 (Living Planet Index[2]). Here we show, however, that this estimate is driven by less than 3% of vertebrate populations; if these extremely declining populations are excluded, the global trend switches to an increase. The sensitivity of global mean trends to outliers suggests that more informative indices are needed. We propose an alternative approach, which identifies clusters of extreme decline (or increase) that differ statistically from the majority of population trends. We show that, of taxonomic–geographic systems in the Living Planet Index, 16 systems contain clusters of extreme decline (comprising around 1% of populations; these extreme declines occur disproportionately in larger animals) and 7 contain extreme increases (around 0.4% of populations). The remaining 98.6% of populations across all systems showed no mean global trend. However, when analysed separately, three systems were declining strongly with high certainty (all in the Indo-Pacific region) and seven were declining strongly but with less certainty (mostly reptile and amphibian groups). Accounting for extreme clusters fundamentally alters the interpretation of global vertebrate trends and should be used to help to prioritize conservation efforts.

Rapid global change is threatening species across the globe[1]. The quantification of biodiversity trends is important to assess whether current investment is slowing or reversing declines, and to identify regions and taxa of concern. Although distilling disparate population trends into a single global index can focus attention on biodiversity trends[2–4], simple metrics can distort the full picture.

Estimates of global biodiversity trends vary depending on their data and mathematical model. The most apocalyptic models gather extensive press coverage, even when based on controversial data (for example, 'biological annihilation'[5], which described trend estimates based largely on expert opinion; or 'insect Armageddon', which is based on data disputed by the original collectors[6]). However, even analyses of the best available data reach conflicting conclusions. An analysis of a global dataset of abundance time series of vertebrates estimated that, on average, vertebrate populations have declined by more than 50% since 1970 (Living Planet Index[2] (LPI)); however, other global analyses found that the mean population size[7,8] and species richness[9,10] have remained stable over similar timeframes. Explanations for the discrepancies have been proposed[8,11–13], but not resolved.

One crucial consideration is that summary indices may be easily misinterpreted. Calculating the geometric mean across populations is the most common and straightforward approach, but is strongly influenced by extremes. To illustrate, imagine an ecosystem in which one population declined by 99%. Even if a second population increased 50-fold or 393 populations increased by 1% (that is, a large net increase), a geometric mean would show a catastrophic 50% decline. Thus, a geometric mean decline of 50% could arise from substantial, widespread loss that is occurring across many populations (we term this the 'catastrophic declines' hypothesis) or from a few extremely declining populations (we term this the 'clustered declines' hypothesis). Both scenarios involve important conservation issues, but suggest vastly different underlying problems and require different mitigation strategies[14], thus distinguishing between them is of real-world importance.

We derive a Bayesian hierarchical mixture (BHM) model to distinguish between the catastrophic and clustered declines hypotheses. The model statistically separates population trends into extreme declines, typical trends and extreme increases (Fig. 1), while accounting for time-series size, within-population fluctuations, number of populations and among-population variance. We test declines in abundance for more than 14,000 vertebrate populations (from the LPI)[15]. We chose LPI data because of its large scope, because the data and analytical details were publicly available, and because previous analyses of these data suggested widespread, global declines[2].

We first examined whether the previous estimate[2] of a mean decline of more than 50% was sensitive to extreme populations: robust declines would support the catastrophic declines hypothesis, whereas high

[1]Department of Biology, McGill University, Montreal, Quebec, Canada. [2]Bieler School of Environment, McGill University, Montreal, Quebec, Canada. [3]Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada. [4]School of Biology and Ecology, University of Maine, Orono, ME, USA. [5]Mitchell Center for Sustainability Solutions, University of Maine, Orono, ME, USA. [6]Centre for Biological Diversity, University of St Andrews, St Andrews, UK. [7]Indicators and Assessments Unit, Institute of Zoology, Zoological Society of London, London, UK. ✉e-mail: brian.leung2@mcgill.ca
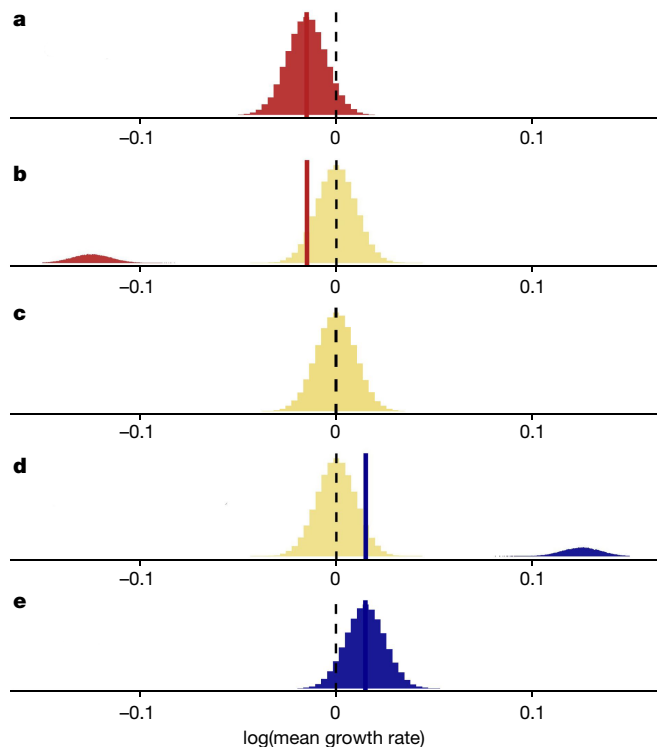
Fig. 1 | Stylized patterns of system-wide growth rates. a–e, Similar geometric mean population growth rates ($\log(N_{t+1}/N_t)$) can reflect contrasting systems. **c**, As a null model, systems can be stable (log-transformed growth rates centred around zero). Deviations can occur in multiple ways. **a**, **b**, Most populations in a system can be in substantial decline (catastrophic declines hypothesis) (**a**) or the system can have multiple clusters, in which the majority of populations show a distribution of growth rates centred around zero but with a small cluster of populations experiencing extreme declines (clustered declines hypothesis) (**b**). Each has the same metric of mean decline (vertical red line indicates a 1.5% annual decline, corresponding to a 50% loss over 50 years), even though most populations in **b** are stable. The converse can also happen; systems in which a small cluster of populations shows an extreme increase, but that show an otherwise stable distribution (**d**) or systems in which most populations increase (**e**) can also occur (vertical blue line indicates a 1.5% annual increase, corresponding to a doubling over 50 years).

sensitivity to a few populations would support the clustered declines hypothesis (Fig. 1). We then applied our BHM model to assess the evidence for catastrophic or clustered declines globally and by region and taxonomy. Finally, we explore two additional conservation issues. First, we test whether declines occur disproportionately in larger animals (large animals tend to have lower reproductive rates), which might release small animals from predation[16]. Second, previous analyses often excluded time series with few data points[10,12,17], but small time series make up most of the available data. We test the effects of their exclusion[18].

Sensitivity of geometric mean to extreme populations

The geometric mean index that underlies the LPI analysis was highly sensitive to extreme populations. Excluding only the 2.4% most-strongly declining populations (354 out of 14,700 populations) reversed the estimate of global vertebrate trends from a loss of more than 50% to a slightly positive growth (Fig. 2). Similarly, excluding 2.4% of the most-strongly increasing populations strengthened the mean decline to 71%. High sensitivity suggests that extreme populations are disproportionately affecting global trend estimates, such that clusters of extreme population decline should be considered explicitly.
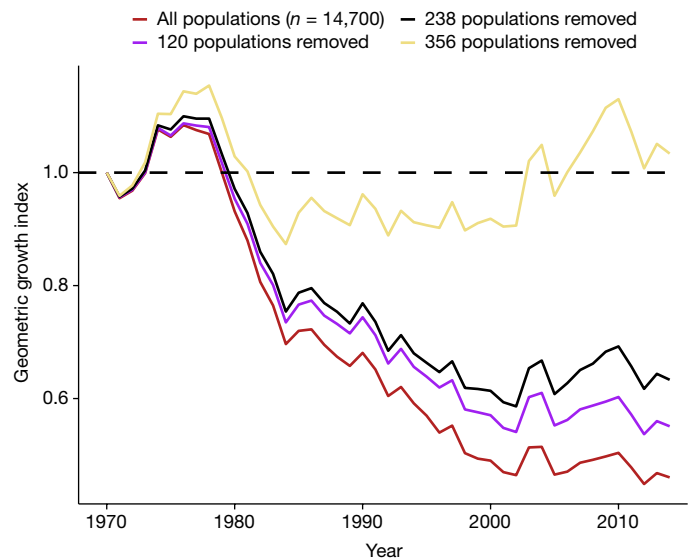

Fig. 2 | Effect of extreme populations on the global growth index. Removing a small fraction of extreme populations strongly influences the geometric growth index, using the LPI dataset. Each line represents a different number of removed populations, ranging from no removals (red line; all 14,700 populations, which show a >50% mean decline) to removing 356 populations (yellow line, the removal of <2.4% of populations switches the global trend from negative to positive). A geometric growth index of 1 indicates no change (dashed horizontal black line).

Evidence for clustered declines

Among the 57 domain–realm–taxon systems of the LPI, 16 systems contained clusters of extreme decline and 8 contained clusters of extreme growth (of those, 3 systems are repeated, as they had both clusters of extreme decline and growth) (Fig. 3 and Supplementary Table 2). Together, clusters of extreme decline accounted for only 1% of populations across systems (2% of populations in the 16 systems in which they occurred). The mean population trend for extremely declining clusters across the 16 systems was $\theta_2 = -3.94$, or approximately 98% loss per year, and deviated substantially from the mean trend of the primary cluster in those systems. Clusters of extreme growth accounted for 0.4% of populations across systems (2.4% in the 8 systems in which they occurred), with $\theta_2 = 3.51$, that is, an explosive 33× growth per year (Fig. 3 and Supplementary Table 2).

Extreme clusters showed some taxonomic and geographic patterns. The largest cluster of extreme declines was in Arctic marine mammals, accounting for 7.6% of populations in that system. However, mammalian systems generally had the fewest clusters of extreme decline (19% of 16 systems), followed by reptile–amphibian systems (21% of 14 systems), whereas bird and fish systems had more clusters of extreme declines (31% of 16 and 45% of 11 systems, respectively) (Fig. 3). Clusters of extreme decline occurred throughout the world, half of which occurred in marine realms, whereas extreme increases occurred more in temperate regions or terrestrial realms (Fig. 3).

Extreme population trends occurred predominantly in small time series. Excluding time series with fewer than 10 points not only removed all but two extreme clusters, but also removed 52% of the data (Supplementary Table 3). The higher frequency of extreme trends among small time series was also apparent in the raw data (Fig. 4). Thus the decision of whether to include small time series will have large effects on the resulting estimates of global trends.

Body size was related to population trends. Larger species had three times more extreme declines than increases (15 compared with 5 clusters of extreme declines compared with extreme increases). Comparatively, smaller species had half as many (8) extremely declining and disproportionately more (7) extremely increasing clusters (Supplementary
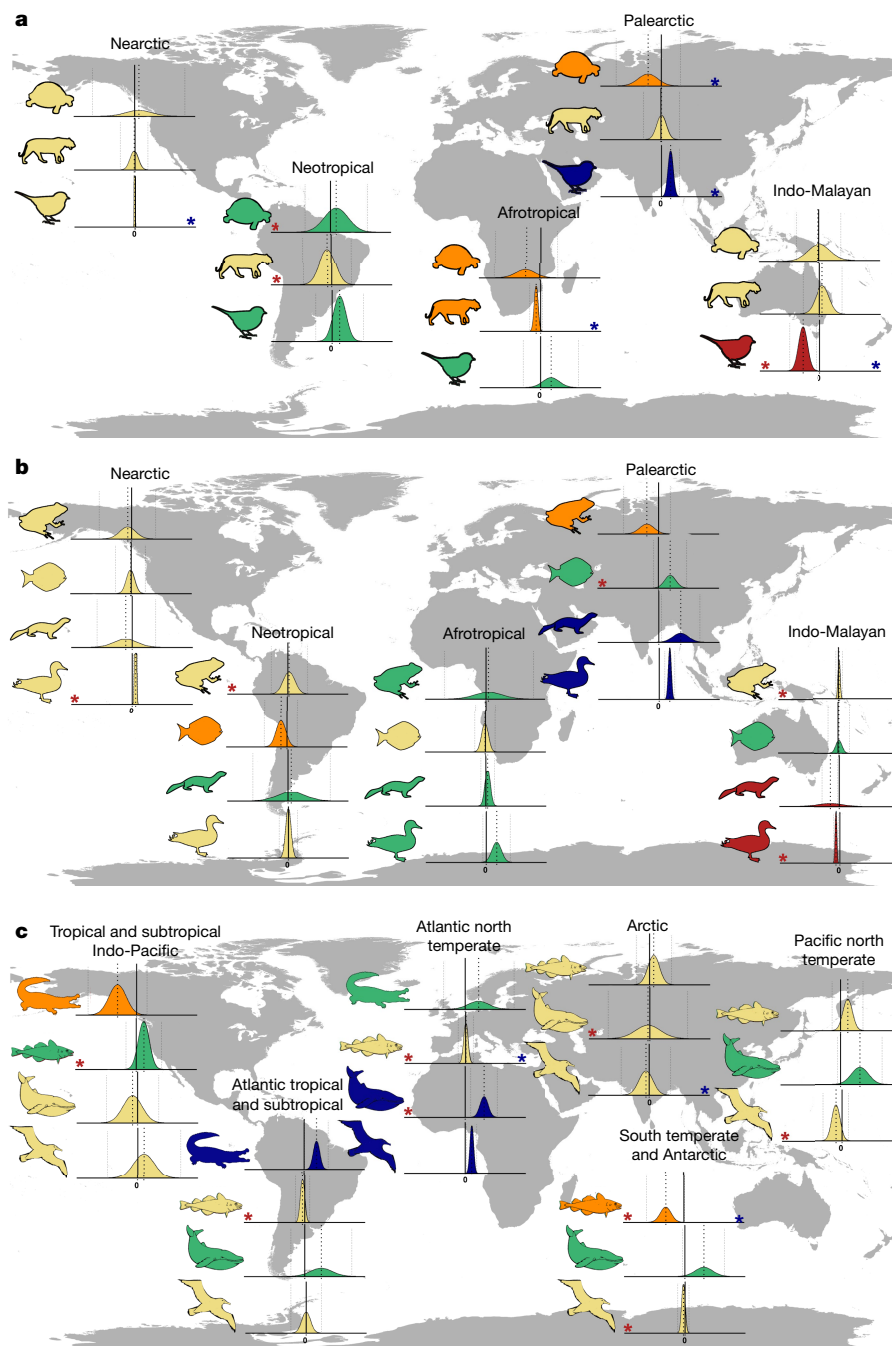
Fig. 3 | Population trends by taxonomic groups and realms. a, The terrestrial realm. **b**, The freshwater realm. **c**, The marine realm. Red and blue asterisks indicate the occurrence of extremely declining clusters (16 systems) and increasing clusters (8 systems), respectively. Distributions show the primary cluster in each system. Red, significant declines; blue, significant increases; orange, strong non-significant declines; green, strong non-significant increases; yellow, weak changes). Maps were created using ArcGIS software by Esri (ArcGIS and ArcMap are the intellectual property of Esri and are used herein under licence. Copyright © Esri. All rights reserved. For more information about Esri software, please visit https://www.esri.com).

Table 4). Although size-specific models included fewer populations, especially for smaller species, the number of clusters was not uniformly lower (as might be expected given a reduction in power); therefore, the differential occurrence of extremely declining versus increasing clusters suggests that large animals are more vulnerable to extreme declines.

Evidence for catastrophic declines

In contrast to the extreme clusters, the primary clusters accounted for the vast majority (98.6%) of populations across the 57 LPI systems. The overall growth rate of primary clusters was close to zero: $\theta_1 = -0.00035$, corresponding to around 1.7% loss over 50 years, given a constant rate across populations and time (Fig. 5). In addition, in contrast to extreme clusters, primary cluster trends were robust to time-series size, as excluding series with fewer than 10 data points yielded a similar overall global trend ($\theta_1 = 0.0043$) (Extended Data Fig. 3).

Although the global BHM model reveals considerably more nuance than a geometric mean index, analysing across systems still masked important patterns. When systems were analysed separately (Supplementary Table 2), primary population clusters were strongly declining
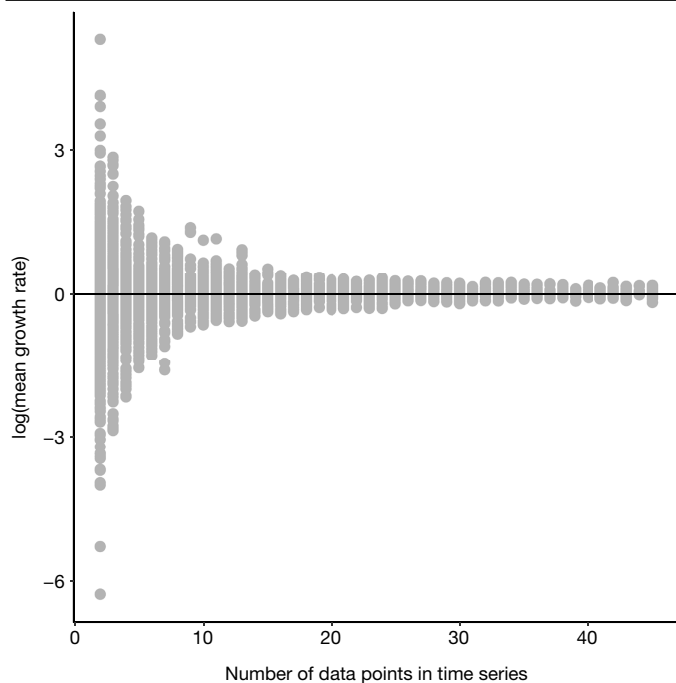
Fig. 4 | Effect of the size of the time series. The number of data points in the time series versus the mean log-transformed value of the geometric mean growth rate.

Fig. 5 | Populations in the primary clusters across all systems, after removal of extreme clusters. The primary cluster of each system is unimodal, but because systems are experiencing decline (or growth) heterogeneously, plotting distributions across systems shows multimodality. Histograms show significantly declining systems (red), strongly but not significantly declining systems (orange) and weak changes or increases (yellow). Vertical lines show thresholds for strongly declining (−0.015) and strongly increasing (+0.015) growth rates, corresponding to an approximate 50% loss or a doubling (over 50 years), respectively. Distributions of primary clusters were calculated based on the mean and s.d. from the hierarchical model, and using the system-specific weights to adjust for species richness.

($\theta_1 < -0.015$) with high certainty (95% credible intervals not overlapping zero) in three systems, all of which occurred in the Indo-Pacific realm (freshwater mammals, freshwater birds and terrestrial birds) (Fig. 3). This suggests that this region has the highest risk of system-wide declines and should be a conservation priority. By contrast, the primary cluster was increasing with high certainty in seven systems, six of which were in temperate regions. In addition, seven additional systems had strongly declining primary population clusters but with less certainty (95% credible intervals overlapped zero), four of which were amphibian or reptile groups. Finally, 14 systems showed strong but low-certainty increases, with no obvious taxonomic nor geographic patterns (Fig. 3).

Each primary cluster also contained variation among populations. In the 10 systems with significant or non-significant mean declines where $\theta_1 < -0.015$, 87% of the individual populations showed strong declines (Fig. 5). These 10 systems accounted for around 20% of the total global vertebrate populations, but for around 61% of strong declines. The multimodality observed in Fig. 5 was an outcome of aggregating unimodal primary clusters across systems, and suggests that there are heterogeneous stressor levels among systems (that is, similar principles to those that cause extreme clusters within systems). The remaining approximately 11% of strongly declining populations were distributed across 47 out of 57 systems; it is unclear whether they represent a deviation from the natural dynamics that are expected to occur in any naturally variable system.

Primary cluster trends were related to body size, but not as predicted. In comparison to the overall patterns for larger animals, the same systems showed significant declines and increases, but two additional temperate systems showed significant increases (Extended Data Fig. 4 and Supplementary Table 4). Smaller species also appeared to decline more than larger species; there were 27 systems in which smaller species had more-negative growth rates than larger species, compared with 18 systems in which the reverse was true. However, analyses of the smaller species were based on substantially fewer populations, and trends were generally not significant (Supplementary Table 4), so patterns remain tentative.
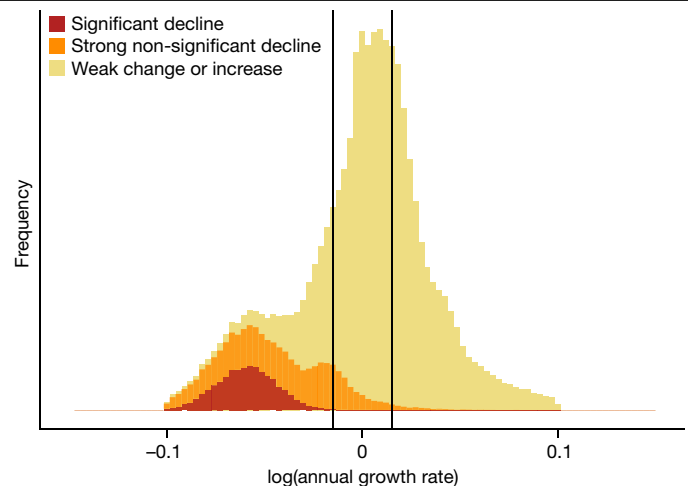
Discussion

By re-analysing a comprehensive dataset of global wildlife population trends, we show that previously estimated global declines are driven by a few extremely declining populations. Removing only 2.4% of declining populations reversed the estimated global trends from more than 50% mean decline since 1970 to a slightly positive growth. Our BHM model revealed that clusters of extreme decline are widespread and occur disproportionately in larger species, and that a few clusters of extreme increase also exist and occur disproportionately in smaller species. This is consistent with previous arguments of 'trophic downgrading'[16].

Clusters of extreme declines were largely due to small time-series datasets. However, neither random sampling error nor 'saw tooth' population dynamics (in which ultimately stable populations experience sudden declines followed by gradual increases) can fully explain this association (see Supplementary Information for a full discussion). Additional explanations are needed. Extreme trends could reflect transient populations that naturally leave or enter a survey area[19], which could represent natural dynamics. Alternatively, researchers may stop sampling after populations become (close to) extirpated, although the converse has also been suggested[20]. A third possibility is that some regions experience both lower sampling effort and greater declines, such that poorly sampled datasets correlate with factors linked to vulnerability, such as lower national wealth or conservation investment. Understanding why small time series contain so many extreme declines is particularly important given that studies that did not find widespread declines often excluded short time series[7,10,12], potentially reconciling divergent findings among studies.

Once extreme clusters were statistically separated, no global trend remained across typical populations (that is, primary clusters; 98.6% of populations). However, aggregating systems into one global trend hid important variation. Three systems, all of which occurred in the Indo-Pacific realm, showed widespread vertebrate declines across typical populations. Moreover, among typical populations smaller species may be faring worse than larger ones. Although these results

were tentative given lower sample sizes and high uncertainty, this trend is contrary to common conservation assumptions and so merits additional research.

Our results emphasize an important point: biodiversity trends within and across regions and taxa are highly disparate. This probably reflects differences in both susceptibility and exposure to anthropogenic environmental change[21–23]. Unravelling this variation is imperative to understand in which regions biodiversity is threatened the most[24] and which conservation actions promote stability or recovery. A productive global conversation about conservation requires that both scientists and media pay more attention to variation and resist the temptation of simple summary indices.

Shifting the message from ubiquitous catastrophe to foci of concern, also touches on human psychology. Continual negative and guilt-ridden messaging can cause despair, denial and inaction[25,26]. If everything is declining everywhere, despite the expansion of conservation measures in recent decades, it would be easy to lose hope. Our results identify not only regions that need urgent action to ameliorate widespread biodiversity declines, but also many systems that appear to be generally stable or improving, and thus provide a reason to hope that our actions can make a difference.

Online content

Any methods, additional references, Nature Research 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/s41586-020-2920-6.

1. IUCN. *The IUCN Red List of Threatened Species*. version 2019-3 http://www.iucnredlist.org (2019).
2. WWF. *Living Planet Report 2018: Aiming Higher* (eds. Grooten, N. & Almond, R. E. A.) (WWF, 2018).
3. Rosenberg, K. V. et al. Decline of the North American avifauna. *Science* **366**, 120–124 (2019).
4. Sánchez-Bayo, F. & Wyckhuys, K. A. G. Worldwide decline of the entomofauna: a review of its drivers. *Biol. Conserv*. **232**, 8–27 (2019).
5. Ceballos, G., Ehrlich, P. R. & Dirzo, R. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl Acad. Sci. USA* **114**, E6089–E6096 (2017).
6. Willig, M. R. et al. Populations are not declining and food webs are not collapsing at the Luquillo Experimental Forest. *Proc. Natl Acad. Sci. USA* **116**, 12143–12144 (2019).
7. 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).
8. Dornelas, M. et al. A balance of winners and losers in the Anthropocene. *Ecol. Lett*. **22**, 847–854 (2019).
9. Vellend, M. et al. Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proc. Natl Acad. Sci. USA* **110**, 19456–19459 (2013).
10. Dornelas, M. et al. Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296–299 (2014).
11. Gonzalez, A. et al. Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology* **97**, 1949–1960 (2016).
12. Leung, B., Greenberg, D. A. & Green, D. M. Trends in mean growth and stability in temperate vertebrate populations. *Divers. Distrib*. **23**, 1372–1380 (2017).
13. McGill, B. J., Dornelas, M., Gotelli, N. J. & Magurran, A. E. Fifteen forms of biodiversity trend in the Anthropocene. *Trends Ecol. Evol*. **30**, 104–113 (2015).
14. Anderson, S. C., Branch, T. A., Cooper, A. B. & Dulvy, N. K. Black-swan events in animal populations. *Proc. Natl Acad. Sci. USA* **114**, 3252–3257 (2017).
15. LPI. *Living Planet Index*. www.livingplanetindex.org/ (2016).
16. Estes, J. A. et al. Trophic downgrading of planet Earth. *Science* **333**, 301–306 (2011).
17. Connors, B. M., Cooper, A. B., Peterman, R. M. & Dulvy, N. K. The false classification of extinction risk in noisy environments. *Proc. R. Soc. Lond. B* **281**, 20132935 (2014).
18. Hanks, E. M., Hooten, M. B. & Baker, F. A. Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecol. Appl*. **21**, 1173–1188 (2011).
19. Youngflesh, C. & Lynch, H. J. Black-swan events: population crashes or temporary emigration? *Proc. Natl Acad. Sci. USA* **114**, E8953–E8954 (2017).
20. Fournier, A. M. V., White, E. R. & Heard, S. B. Site-selection bias and apparent population declines in long-term studies. *Conserv. Biol*. **33**, 1370–1379 (2019).
21. Newbold, T. et al. Ecological traits affect the response of tropical forest bird species to land-use intensity. *Proc. R. Soc. Lond. B* **280**, 20122131 (2013).
22. Venter, O. et al. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun*. **7**, 12558 (2016).
23. Allan, J. R. et al. Hotspots of human impact on threatened terrestrial vertebrates. *PLoS Biol*. **17**, e3000158 (2019).
24. Blowes, S. A. et al. The geography of biodiversity change in marine and terrestrial assemblages. *Science* **366**, 339–345 (2019).
25. O'Neill, S. & Nicholson-Cole, S. "Fear won't do it": promoting positive engagement with climate change through visual and iconic representations *Sci. Commun*. **30**, 355–379 (2009).
26. Brennan, L. & Binney, W. Fear, guilt, and shame appeals in social marketing. *J. Bus. Res*. **63**, 140–146 (2010).

Article

Methods

Dataset

The publically available LPI dataset includes 15,241 vertebrate populations from 3,510 species[15]. When a species contained both finer-resolution estimates within a country (2,593 entries) and a country-wide aggregate, we excluded the country-wide aggregate (537 entries), yielding 14,700 populations. LPI groups species into 57 systems defined by a combination of habitat domain (terrestrial, freshwater or marine), biogeographical realm (terrestrial/freshwater realms, Afrotropical, Nearctic, Neotropical, Palearctic, Indo-Pacific; marine, Arctic, Atlantic north temperate, Atlantic tropical/sub-tropical, Pacific north temperate, Indo-Pacific tropical/sub-tropical, South-temperate/Antarctic) and taxonomic grouping (fish, Actinopterygii, Elasmobranchii, Holocephali, Myxini, Chondrichthyes, Sarcopterygii, Cephalaspidomorphi; birds, Aves; mammals, Mammalia; herps, Amphibia, Reptilia) (Extended Data Figs. 5–8).

To analyse the effect of body size, we obtained information on each taxonomic group. Given the diversity of vertebrate groups in this dataset and the different conventions across groups, we used different measures of body size for each taxonomic class on the basis of data availability. For birds ($n = 1,397$), mammals ($n = 534$) and reptiles (Squamata, $n = 132$; Testudines, $n = 44$; and Crocodylia, $n = 16$) we used estimates of the mass of the species (in grams) collated in an extensive comparative dataset[27]. When mass data were missing for a species ($n = 14$ birds; $n = 1$ mammal; $n = 25$ reptiles), we estimated body mass as the geometric mean of available mass estimates for species in that genus. For fishes (Chondrichthyes, Osteichthyes and Agnatha; $n = 1,211$), estimates of mass were scarce for most species, so we instead used estimates of total length or standard length (in centimetres), both of which were extracted from FishBase[28] using the rfishbase R package[29]. These length estimates are an imperfect proxy for size (in terms of mass) given the variability in body plans across groups, but given the large amount of variation across these groups it suffices as a way to broadly categorize species into distinct size classes. For amphibians, we used estimates of snout–vent length (in millimetres) as our proxy for body size, as this is the most widely available metric of size across species. Data on snout–vent length for amphibian species ($n = 175$) were extracted from a comprehensive ecological trait dataset: AmphiBio[30].

Sensitivity of the geometric indices to extreme population trends

The LPI analysis was based on a geometric mean approach, calculated by summing across log-transformed growth rates[31]. We recreated the geometric-mean-based analyses (see Supplementary Information 1a for full details and model formulation) and examined the sensitivity of the global estimate to extreme populations. We ordered populations and sequentially removed the largest observed decline, determining the effect of each removal on the global estimate of biodiversity loss. Low sensitivity would indicate that many or most populations are declining, supporting the catastrophic declines hypothesis. High sensitivity—that is, if removal of relatively few populations switched the strongly negative global trend to neutral or positive—would support the clustered declines hypothesis. For balance, we also examined sensitivity to sequential removal of the greatest increasing populations.

Catastrophic versus clustered declines approach

We developed an approach to separate extreme population clusters the growth or decline of which statistically deviated from typical population trends, such that a small number of extreme populations would no longer mask trends of the majority of populations (Fig. 1). Although some summarization is needed to understand global trends, heterogeneous growth rates and potentially multimodal distributions could be expected, given multiple stressors with diverse effects, and differences in species vulnerabilities. We used a BHM model as our statistical architecture, as it has several desirable properties: (1) it can represent the null model and assess deviations from it; (2) it enables testing for both negative and positive extremes (sometimes both existed in the same system); (3) it quantifies the magnitude and proportion of those extremes; (4) it provides a coherent way to separate extreme populations from the majority of populations (the primary cluster), which enables tests of the clustered and catastrophic declines hypotheses; (5) it provides a measure of uncertainty as a direct outcome of analysis (through the posterior distribution); and (6) it accounts for population fluctuations and adjusts for the number of data points in the time series.

First, we specify the null model. Even in a system with no overall trend, we expect stochastic fluctuations in population size. We also expect some populations to be increasing or decreasing during any time interval, given complex, real-world ecological dynamics. Thus, the null model should include among-population heterogeneity, and therefore consists of a distribution of growth rates (Fig. 1c). Statistical deviations from this null model could be caused by a shift in the overall distribution, in which a system-wide mean growth <0 (that is, decline) could indicate a risk to the entire system, which would support the catastrophic declines hypothesis (Fig. 1a). Alternatively, statistical deviation from the null model could be caused by a few populations that experience extreme declines, which is consistent with the clustered declines hypothesis (Fig. 1b).

To specify our model, we begin with a standard Bayesian hierarchical formulation (that is, it does not yet contain mixtures of distributions). We define θ and τ as the system-wide mean and variance, respectively, of log-transformed growth rates across all populations in the system (that is, hyperparameters in Bayesian terminology). θ and τ determine the distribution of the log-transformed population trends (μ_i) and define the properties of the overall system. However, within-population dynamics are also occurring, and the log-transformed growth rates for population i at time t are modelled as a population trend (μ_i) and within-population fluctuations (σ) (see Supplementary Information 1b for full details and model formulation).

Using a standard Bayesian hierarchical model, we can test the catastrophic declines hypothesis by determining the probability that a system-wide mean value of $\theta < 0$. Testing the clustered declines hypothesis, however, requires a mixture model to assess the evidence for the occurrence of clusters. Thus, we define K as the number of clusters in the mixture, f_k is the fraction of populations in the kth cluster, and θ, τ and f denote the vectors of the parameters for the K clusters.

To test the clustered declines hypothesis, we modelled three clusters: a primary cluster, corresponding to the typical trend; a negative extreme cluster; and a positive extreme cluster (Fig. 1). Although our main interest was in the mechanisms behind apparent global population declines (that is, catastrophic versus clustered declines hypotheses), we also assayed positive extreme clusters so that analyses were not biased to find only negative population trends. We considered four cluster combinations: (1) a single distribution; (2) a primary distribution and a negative extreme distribution; (3) a primary distribution and a positive extreme distribution; or (4) a primary distribution and both positive and negative extreme distributions (Fig. 1). For referencing purposes, we denote $k = 1$ as the primary cluster, $k = 2$ as the negative extreme cluster, and $k = 3$ as the positive extreme cluster. Reality need not be bi-modal (or tri-modal), but exploring generalities in trends necessitates some aggregation. Nonetheless, the extreme clusters identified by the mixture model could contain multiple extreme modes in the data (or even result from a skewed distribution). With any of these deviations, model selection would still choose the mixture model as explaining the data better than a single normal distribution (see Supplementary Information 1c for full details and model formulation).

We used the (lowest) deviance information criterion value to select the mixture model with the strongest statistical evidence[32]. The catastrophic declines hypothesis would be supported by a mean decline of the primary population cluster ($\theta_1 < 0$ and credible intervals did not

overlap zero), and would be particularly severe if the mean θ_1 was also strongly negative (for example, $\theta_1 = -0.015$ would correspond to >50% loss over 50 years). The clustered declines hypothesis would be supported if the deviance information criterion selected a mixture with a negative extreme cluster (combinations 2 or 4 above). The catastrophic and clustered declines hypotheses are not mutually exclusive, as a system could have both a negative extreme cluster and declining primary cluster. A large fraction of populations in the negative extreme cluster (f_2) could also be interpreted as widespread catastrophic declines, but this did not occur in our results. Although our hypotheses focus on understanding declining trends, our model will also detect increases in abundances.

To estimate the model parameters, we used Bayesian analyses and the Markov chain Monte Carlo algorithm, which simultaneously estimated uncertainty. For each Bayesian analysis, we ran 3 chains, each with 10,000 iterations (3,000 used for burn-in). Convergence was determined using $\hat{R} \approx 1$. Values for all parameters across all systems ranged from ($0.999 < \hat{R} < 1.005$). Bayesian analyses were conducted using the STAN language[33], and processed and analysed in R[34].

Additionally, we explored the theoretical behaviour of each model, including the geometric mean model, in the presence of clustered declines (Supplementary Information 1d, 2a), and our catastrophic and clustered declines approach given our selection of priors, application of constraints and other modelling choices; these simulation analyses showed that our approach yielded appropriate theoretical behaviour (Extended Data Fig. 1 and Supplementary Information 1e, 2b). Finally, we conducted sensitivity analyses and showed that results were robust to modelling choices (Extended Data Fig. 2, Supplementary Information 2c and Supplementary Table 1).

Application of the catastrophic and clustered approach to LPI data

We tested for extreme clusters in each of the 57 domain–realm–taxon systems of the LPI, by choosing the mixture model with the lowest deviance information criterion value. We also examined the number of populations in each cluster, as a fraction of the total number of populations, scaled using LPI system-specific weightings[35] (see Supplementary Information 1f for more details).

Next, we examined evidence for the catastrophic declines hypothesis in each system by searching for negative mean growth rates in the primary cluster (θ_1). We defined 'high certainty' of decline (or increase) as 95% credible intervals that did not overlap zero, and 'strong' decline as $\theta_1 < -0.015$, corresponding to a ~50% decline if it persisted for 50 years ($\theta_1 > 0.015$ was used for a strong positive relations, corresponding to a doubling over 50 years).

We assessed the effect of small time series on both extreme clusters and trends in primary clusters, by omitting all data with fewer than 10 points, as has often been done in other studies[12]. These small time series accounted for 52% of the population estimates (7,110 populations remained in the analysis).

Finally, we examined whether trends differed between large- versus small-bodied animals. Within each class (but with Agnatha lumped with Osteichthyes), we scaled body size as standard deviations on the natural log scale—thus creating an index of relative species size within a taxonomic group. In two cases, we separated out different groups within a class that had relatively distinct body plans that would influence this size scaling. We scaled size within the superorder Batoidea (Rajiformes, Myliobatiformes and Torpediniformes) and separately scaled size for the rest of the chondrichthyans (Selachimorpha and Holocephali). For the amphibians, we separated out the orders Caudata and Anura and scaled size within each of these groups. For each taxonomic group, we scaled body size and separated species into larger-than-average (hereafter 'larger') versus smaller-than-average (hereafter 'smaller') species. This yielded 9,596 populations from 1,765 larger species, and 5,103 populations from 1,745 smaller species. We then reran the BHM model for larger animals and again for smaller animals. Body sizes were divided unevenly among habitat domains and realms; 12 domain–realm–taxon systems contained ≤1 smaller species so were excluded from the small-animal model.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
Data can be obtained from the LPI database (www.livingplanetindex.org), AmphiBio[30] (https://figshare.com/articles/Oliveira_et_al_AmphiBIO_v1/4644424), FishBase (www.fishbase.org)[28] and life-history traits can be obtained from the amniote life-history database[27] (https://doi.org/10.6084/m9.figshare.c.3308127.v1).

Code availability
Code for the BHM model is available at: https://doi.org/10.5281/zenodo.3901586.

27. Myhrvold, N. P. et al. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology* **96**, 3109 (2015).
28. Froese, R. & Pauly, D. *FishBase* version 12/2019 www.fishbase.org (2019).
29. Boettiger, C., Lang, D. T. & Wainwright, P. C. rfishbase: exploring, manipulating and visualizing FishBase data from R. *J. Fish Biol.* **81**, 2030–2039 (2012).
30. Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C. & Costa, G. C. AmphiBIO, a global database for amphibian ecological traits. *Sci. Data* **4**, 170123 (2017).
31. Collen, B. et al. Monitoring change in vertebrate abundance: the living planet index. *Conserv. Biol.* **23**, 317–327 (2009).
32. Gelman, A., Hwang, J. & Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**, 997–1016 (2014).
33. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
34. R Core Team. *R: A Language and Environment for Statistical Computing* http://www.R-project.org/ (R Foundation for Statistical Computing, 2016).
35. McRae, L., Deinet, S. & Freeman, R. The diversity-weighted living planet index: controlling for taxonomic bias in a global biodiversity indicator. *PLoS ONE* **12**, e0169156 (2017).

Author contributions Authors are listed in order of their contributions. B.L. formulated the BHM model, conducted analyses and wrote the majority of the paper. A.L.H. discussed and clarified the ideas and had a central role in the writing of the paper. D.A.G. discussed and clarified the ideas, synthesized the data and contributed to the writing of the paper. B.M. discussed and clarified the ideas, and commented on the manuscript. M.D. discussed and clarified the ideas, commented on and improved the presentation of the manuscript. R.F. discussed and clarified the ideas, and provided insight into the LPI data and analyses.
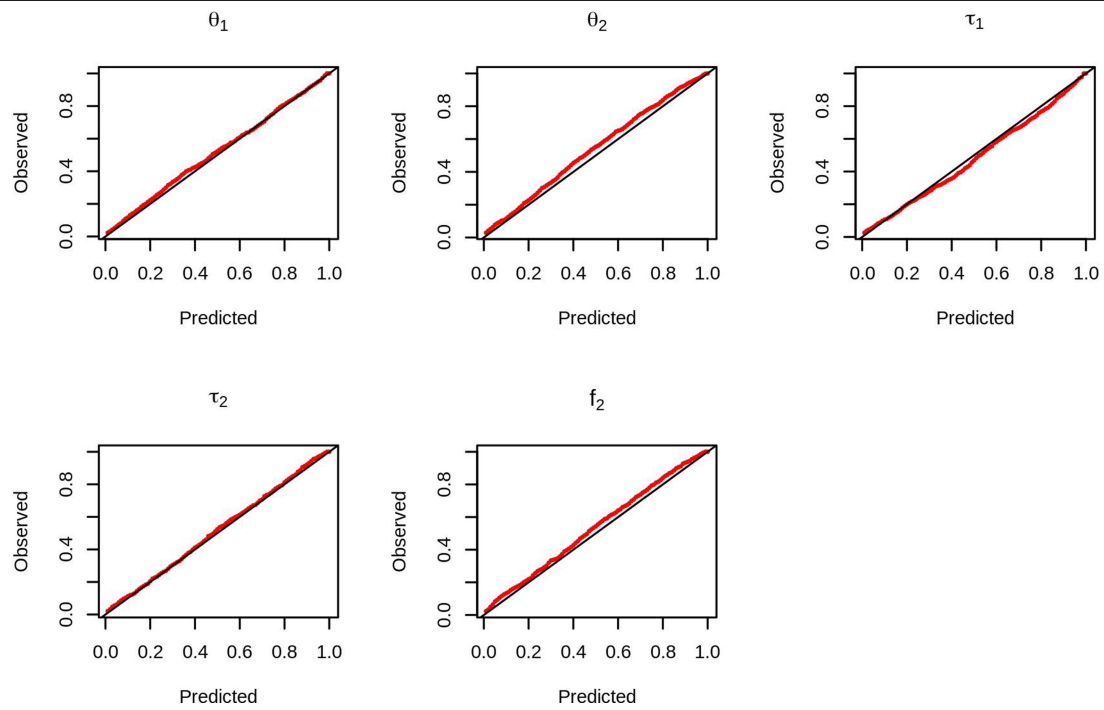
Extended Data Fig. 1 | Theoretical analyses of BHM model. The p–p plots show that the posterior distributions for each estimated parameter are unbiased and largely follow a 1:1 line for each hyper parameter (σ, τ) as well as the fraction in each cluster (f_1, $f_2 = 1 - f_1$). The 1:1 line is the theoretic expectation, indicating that the true parameter value falls below the 0.01 quantile 1% of the time, the 0.02 quantile 2% of the time, and so on.

Sensitivity of main cluster growth rate to threshold value

0.5*sd threshold

Growth rates, 1*sd threshold
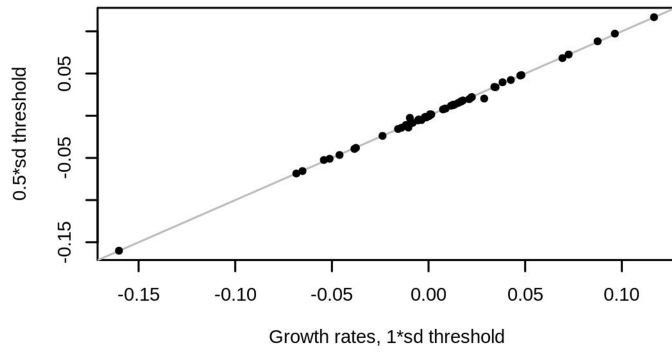
Sensitivity of main cluster growth rate to offset value

Offset = 5% of $\bar{x}$

Growth rates, offset = 1% of $\bar{x}$
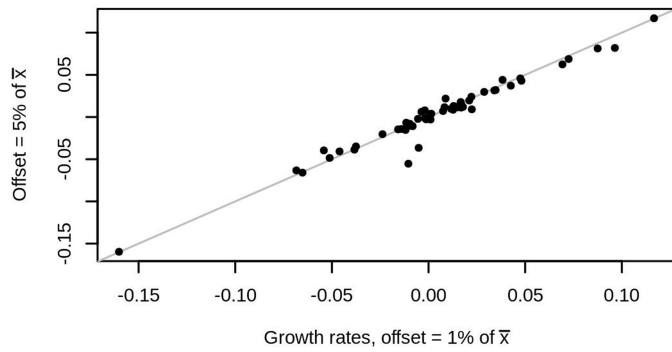
Extended Data Fig. 2 | Sensitivity analyses of primary cluster trends. The trends of the primary clusters (θ_1), for the main analysis (x axis) versus the sensitivity analysis (y axis) for the threshold for extreme clusters (top) and the offset when $n = 0$ was observed (bottom).
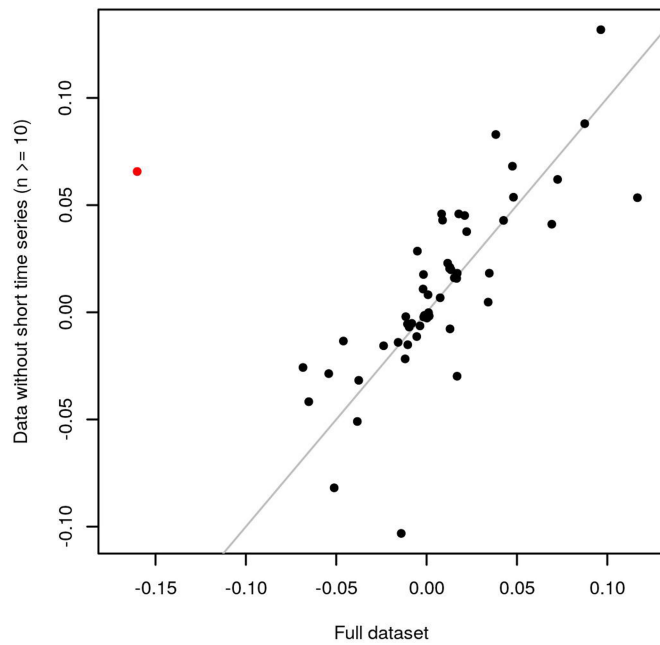
Extended Data Fig. 3 | Effect of small time series on primary cluster trends.
Each point represents a trend estimate for the primary cluster of a system, with
the full dataset (*x* axis) versus data excluding time series with less than 10 data
points (*y* axis). The red dot indicates the freshwater Indo-Pacific mammals,
which was reduced from 22 populations (full) to 2 populations (only data with
at least 10 data points).
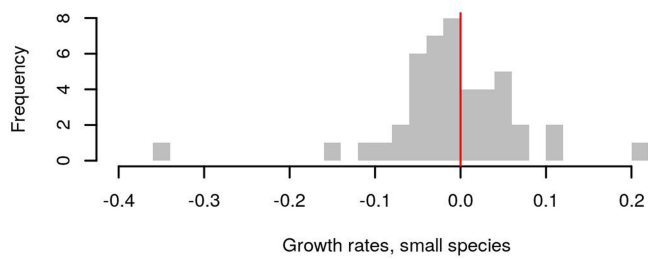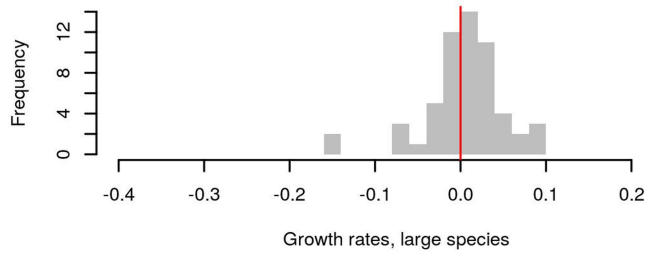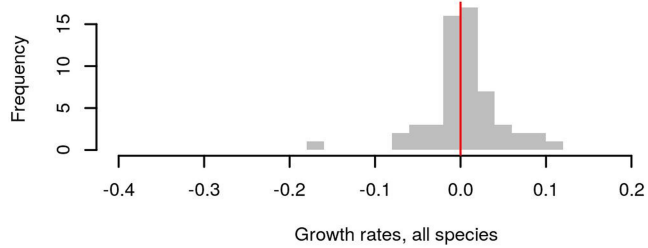
Extended Data Fig. 4 | Mean trends of primary clusters across systems calculated using the BHM model. Top, all species (14,700 populations). Middle, only large species (9,596 populations). Bottom, only small species (5,103 populations). The small species appear to be declining more than large species, although this finding needs to be interpreted with caution, as most primary distributions did not significantly deviate from zero for small species.
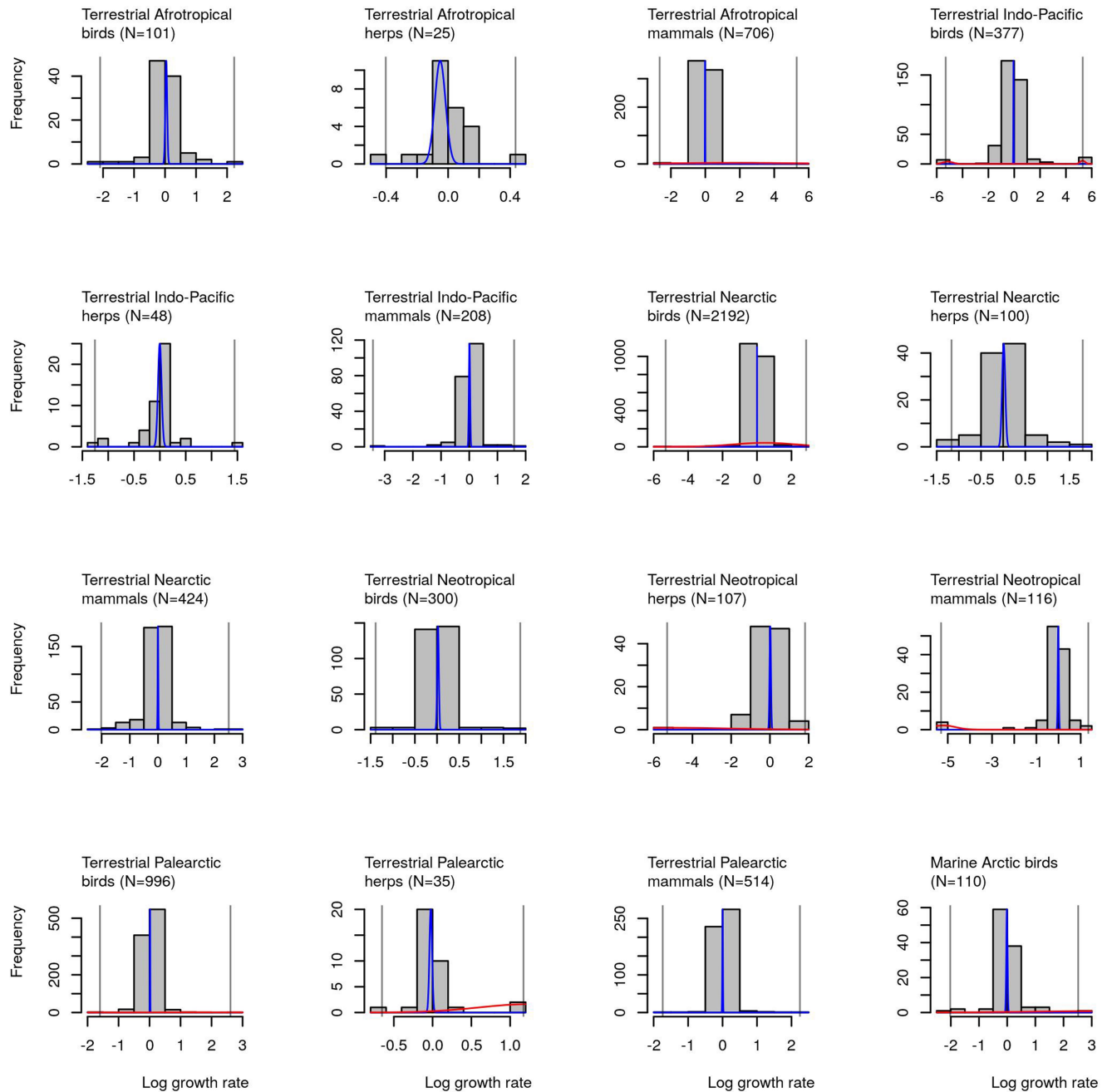
Extended Data Fig. 5 | Histograms of observed growth rates and output of the BHM model for systems 1–16. Blue line, primary cluster; red line, extreme cluster(s) from the model. Grey vertical lines show the range of observed values. In comparing the model output to the data we show the following. (1) The variation of the BHM primary cluster (blue line) is much lower than the raw data, because the BHM separates variation in among-population trends from variation due to within-population fluctuations. (2) The BHM model identifies evidence for extreme clusters in both directions (for example, terrestrial Indo-Pacific birds) or only one direction (for example, terrestrial Neotropical mammals), but not for other apparent clusters (for example, terrestrial Indo-Pacific herps). The BHM integrates the magnitude of within-population fluctuations, time-series sizes, number of populations, among-population variance, and the magnitude and frequency of the extreme populations in determining whether additional (extreme) clusters are needed to account for the observations.
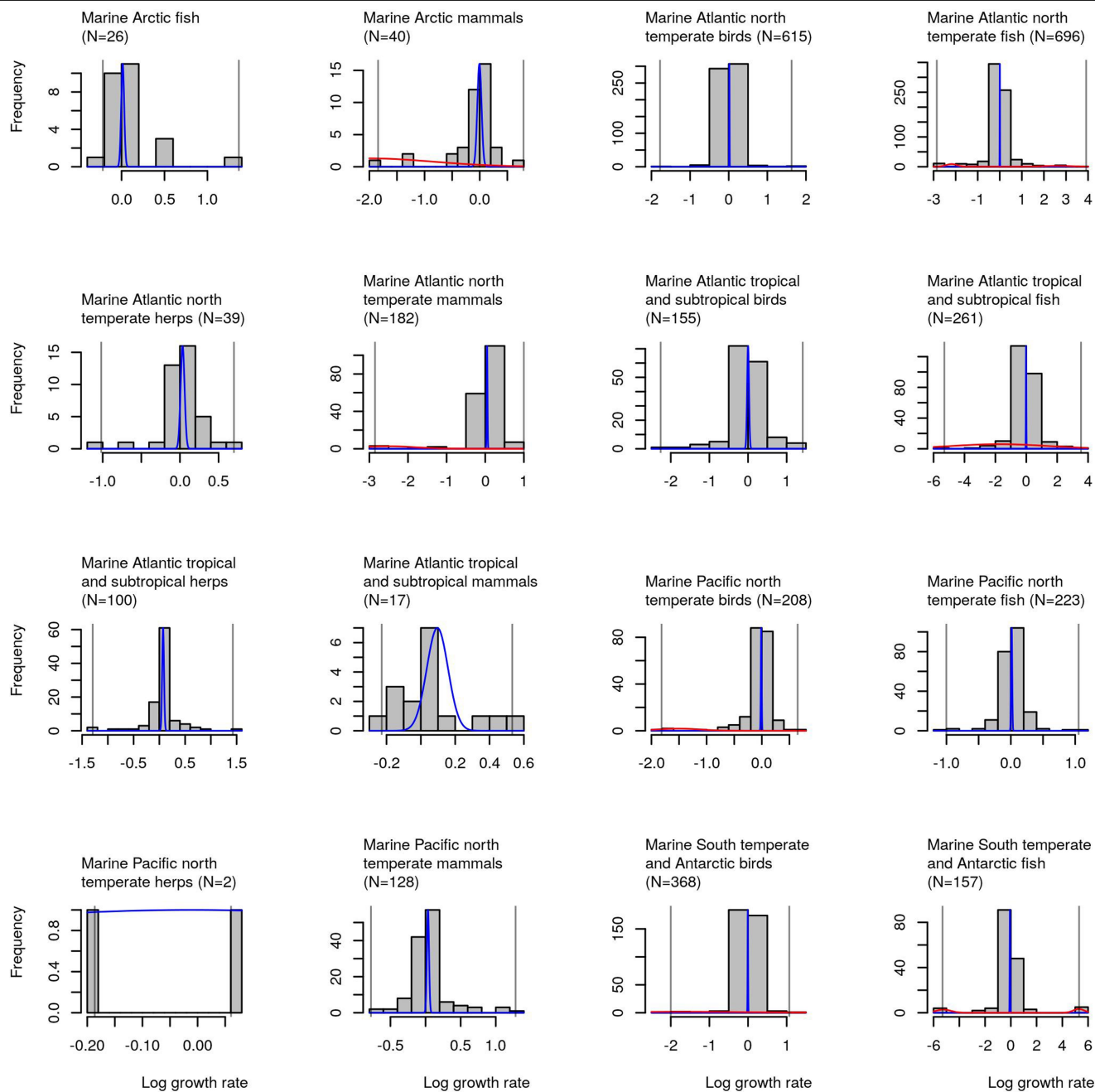
Extended Data Fig. 6 | Histograms of observed growth rates and output of the BHM model for systems 17–32. Blue line, primary cluster; red line, extreme cluster(s) from the model. Grey vertical lines show the range of observed values. For further information, see Extended Data Fig. 5.
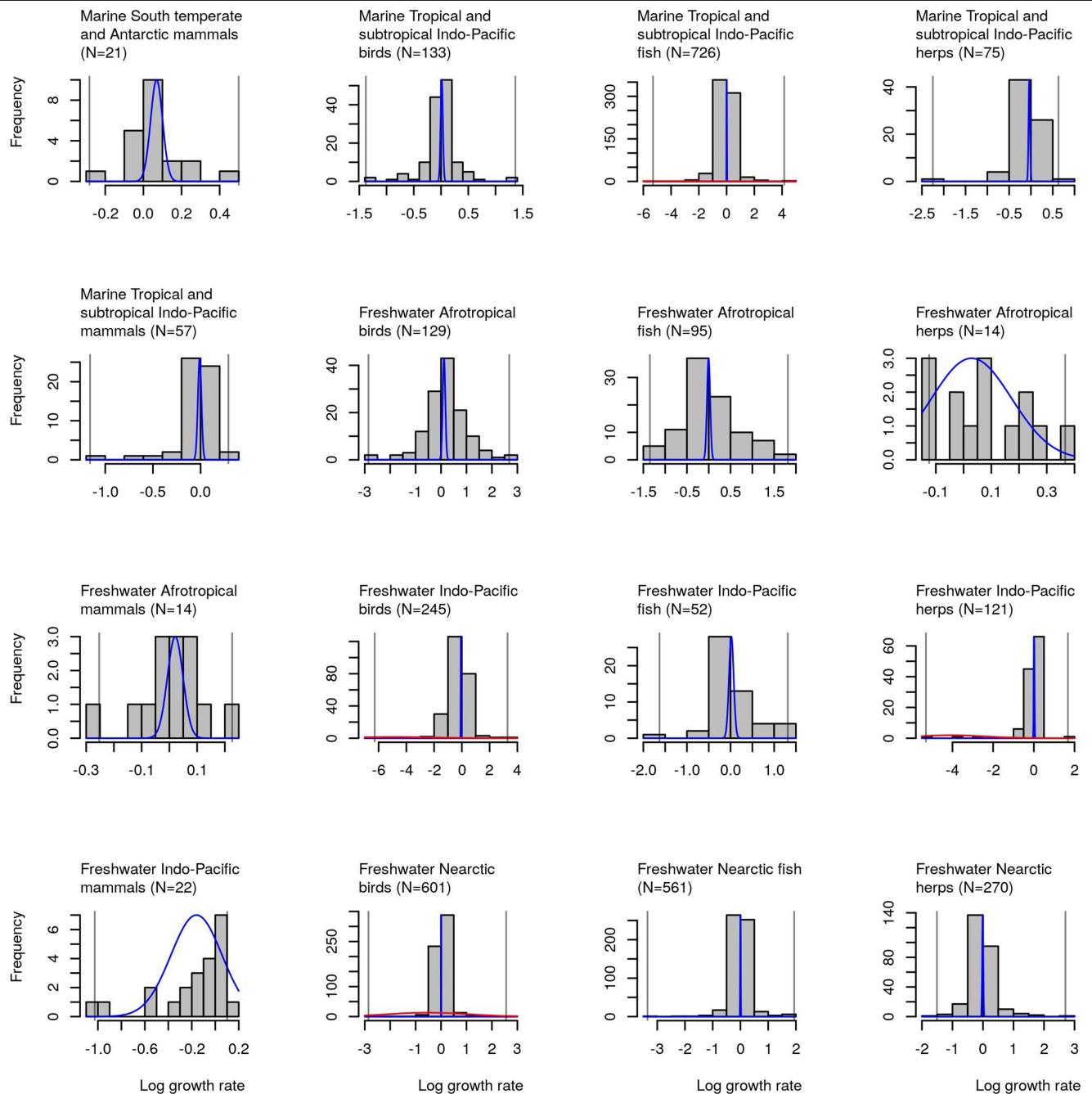
Extended Data Fig. 7 | Histograms of observed growth rates and output of the BHM model for systems 33–48. Blue line, primary cluster; red line, extreme cluster(s) from the model. Grey vertical lines show the range of observed values. For further information, see Extended Data Fig. 5.
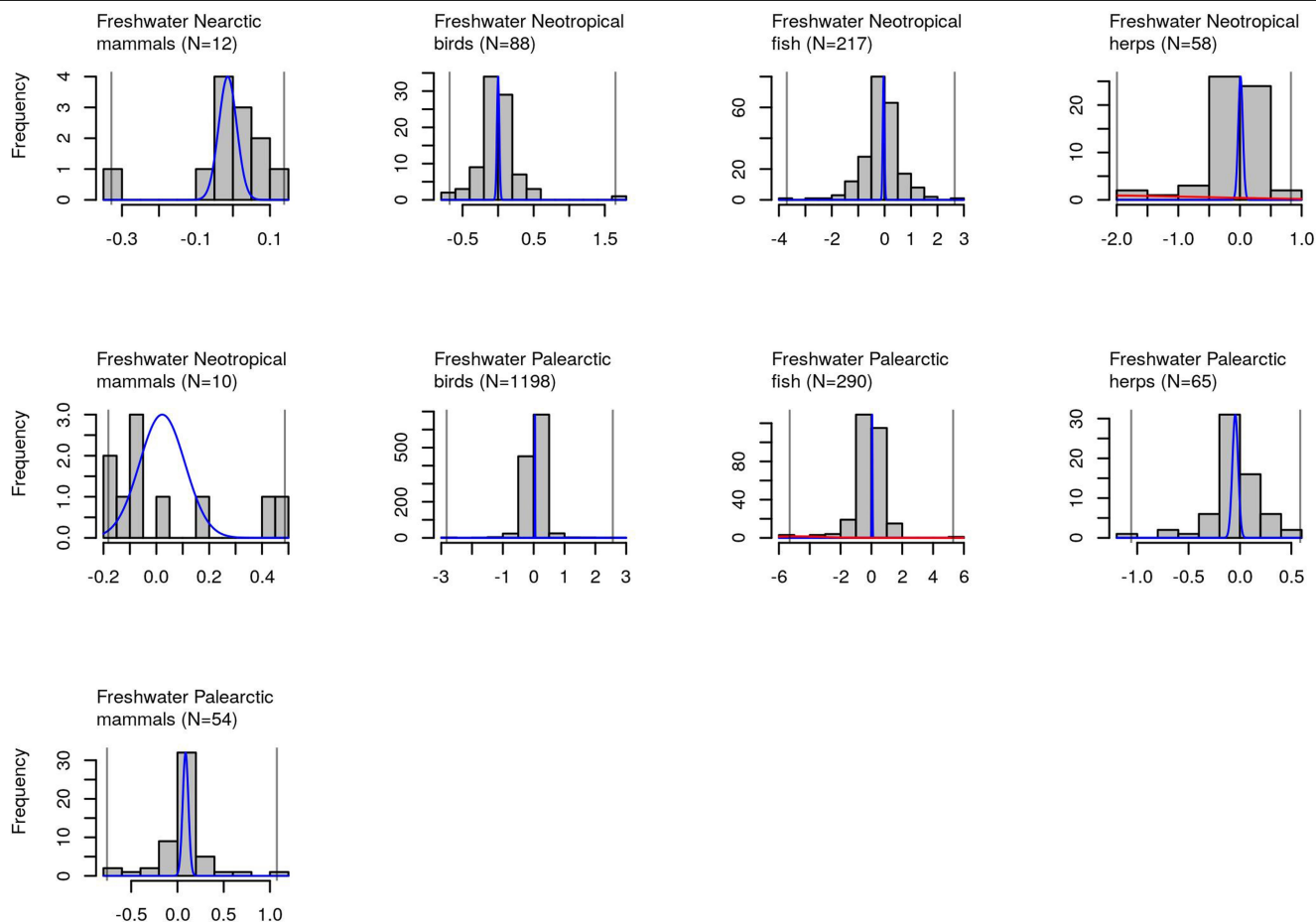
Extended Data Fig. 8 | Histograms of observed growth rates and output of the BHM model for systems 49–57. Blue line, primary cluster; red line, extreme cluster(s) from the model. Grey vertical lines show the range of observed values. For further information, see Extended Data Fig. 5.

nature research

Corresponding author(s): Brian Leung

Last updated by author(s): Aug 27, 2020

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	Data from the Living Planet Index database. <www.livingplanetindex.org/>. (2016) was scraped in R 3.6.3. Data was extracted from Fishbase using rfishbase 3.0.4.
Data analysis	Bayesian analyses were conducted using the STAN 2.14 language, and processed and analyzed in R 3.6.3. The lme4 1.1-23 package was referenced in the text. Custom code from this article can be obtained at: https://doi.org/10.5281/zenodo.3901586

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Data can be obtained from the Living Planet Index database. <www.livingplanetindex.org/>. (2016), AmphiBIO database from <https://figshare.com/articles/Oliveira_et_al_AmphiBIO_v1/4644424>, Fishbase database <www.fishbase.org>, and mammal, bird and reptile life history traits from <https://doi.org/10.6084/m9.figshare.c.3308127.v1>

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	The study re-examined previous estimates of global vertebrate declines from the Living Planet Index (LPI), and demonstrates that it is driven by a small fraction of extreme population trends. The study then goes on to use a Bayesian Hierarchical Mixture Model to separate the extreme clusters and primary cluster. The LPI dataset that was publicly available consists of 15241 vertebrate populations from 3510 species, grouped into 57 domain-realm-taxon systems (hereafter just 'systems') (LPI 2016). Each system was analyzed separately, and also was combined into a global metric using weighting factors (adjusted for species richness in each system) from LPI, for comparability.
Research sample	The data was obtained from the Living Planet Index database. <www.livingplanetindex.org/>. (2016), and consisted of 15241 vertebrate populations. To avoid double counting, when a species contained both finer resolution estimates within a country (2593 entries) as well as a country-wide aggregate, we excluded the country-wide aggregate (537 entries). This resulted in 14700 populations remaining in our analysis. Each system was defined by a combination of habitat domain (terrestrial, freshwater and marine), biogeographic realm, and taxonomic grouping (Fish=Actinopterygii, Elasmobranchii, Holocephali, Myxini, Chondrichthyes, Sarcopterygii, Cephalaspidomorphi; Birds=Aves, Mammals=Mammalia, Herps = Amphibia, Reptilia). Terrestrial and freshwater habitat domains were separated into five realms (Afrotropical, Nearctic, Neotropical, Palearctic, and Indo-Pacific), whereas the marine domain was separated into six realms (Arctic, Atlantic north temperate, Atlantic tropical/sub-tropical, Pacific north temperate, Indo-Pacific tropical/sub-tropical, and South-temperate/Antarctic).
Sampling strategy	All population time-series data in the LPI dataset were used. To avoid double counting, when a species contained both finer resolution estimates within a country (2593 entries) as well as a country-wide aggregate, we excluded the country-wide aggregate (537 entries). This resulted in 14700 populations remaining in our analysis.
Data collection	The data was obtained by Dan Greenberg, and downloaded from publicly available databases identified in the data availability statement
Timing and spatial scale	Data were analyzed from 1970-2014, as these coincided with the analyses from the Living Planet Index. The spatial scale for the analysis was global. The data was comprised of 14700 populations across many studies, and thus was measured at many scales. Thus, relative changes per population was used.
Data exclusions	To avoid double counting, when a species contained both finer resolution estimates within a country (2593 entries) as well as a country-wide aggregate, we excluded the country-wide aggregate (537 entries). This resulted in 14700 populations remaining in our analysis.
Reproducibility	This is not relevant, as the existing LPI database was used. The purpose of the study was not an experiment, but instead to re-analyze the available information on vertebrate trends, to evaluate whether previous estimates of decline (>50%) were due to clusters of extremely declining populations, and to separate and analyze extreme clusters and primary clusters separately.
Randomization	This is not relevant, as the existing LPI database was used, chosen for its impressive size and geographic coverage, and because previous analyses of these data suggested broad-scale average vertebrate.
Blinding	This is not relevant. Blinding as done in clinical trials, where group assignments of individuals is hidden from some researchers. Primary data collection and experiments were not conducted in this study.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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