Assessing the sensitivity and robustness of the Living Planet Index through simulated population dynamics: strengths, stability, and challenges

Cristian A. Cruz-Rodríguez[1,2] , Gaëlle Mével[3] , Shuaishuai Li[2,4] Robin Freeman[5,6], Timothée Poisot[1,2], Valentina Marconi[5,6], Louise McRae[5,6], Jessica Currie[7], Janaína A. Serrano[2,4] , Maria Isabel Arce-Plata[1,2], Sarah Ravoth[8], Philippa Oppenheimer[5,9], Maximiliane Jousse[2,4], Sandra Emry[6] and David AGA Hunt[4]

Corresponding author: Cristian A. Cruz-Rodríguez cruzrodriguezcristian@gmail.com

[1] Université de Montréal, Biological Sciences, 1375 Ave. Thérèse-Lavoie-Roux, Montréal, CA, H2V 0B3.

[2] Quebec Centre for Biodiversity Science, 1205 Dr. Penfield Ave, Montreal, CA, H3A 1B1.

[3] McGill University, Department of Geography, 805 Sherbrooke St W, Montreal, CA, H3A 0B9.

[4] McGill University, Department of Biology, 1205 Dr. Penfield Ave, Montreal, CA, H3A 1B1.

[5] Institute of Zoology, Zoological Society of London, Outer Circle, London, UK, NW1 4RY.

[6] University College London, Faculty of Life Sciences, Division of Biosciences, Department of Genetics, Evolution and Environment (GEE), Research, Gower St, Bloomsbury, London, UK, WC1E 6BT.

[7] WWF-Canada, 410 Adelaide Street West, Toronto, CA, M5V 1S8.

[8] University of British Columbia, Department of Zoology, 6270 University Blvd 4200, CA, V6T 1Z4.

[9] University of Reading, School of Biological Sciences, Ecology and Evolutionary Biology, Reading, UK, RG6 6EX.

Abstract

Understanding population change through time is crucial for effective conservation of biodiversity. The Living Planet Index (LPI) is a key indicator for tracking global species abundance trends under the Kunming-Montreal Global Biodiversity Framework, offering a picture of population change over time. However, the sensitivity of the index to population abundance being recorded as absent in a given year (zero values) despite abundances being recorded in prior and following years or no recording taking place in a given year (missing values) influencing the index and associated trends has not been fully explored. Using simulations, we assessed how missing data and zero values influence the index and associated trends. We found that the LPI method is informative with complete time series. Missing data raises variation but only results in deviations from baseline trends. In contrast, the presence and distribution of zeros within the underlying data substantially influence the results and produce significantly lower trends. These findings highlight the need to develop complementary approaches for evaluating how data heterogeneity influences LPI trends, towards a comprehensive understanding and better-informed decisions on biodiversity change.

1. Introduction

The Living Planet Index (LPI) is considered a measure of the state of the world's biodiversity[1]. It is based on combined population trends across vertebrate species[2] and has become a leading worldwide indicator to discuss global tendencies in nature towards international biodiversity goals (CBD/COP/DEC/15/4, CBD/COP/DEC/15/5). The LPI uses the largest repository of temporal vertebrate population abundance data, the *Living Planet Database* (LPD). This database contains vertebrate population datasets extracted from scientific papers, online databases, government sources, and expert reports[2–4]. Despite thorough processing, the data remains heterogeneous due to significant differences between taxonomic groups in global population time series, including variations in geographical and temporal resolution[5,6] as well as differences in the extension of the time series[5–7]. Additionally, many series contain missing data, zeros, and irregular or sparse sampling intervals [5] that contribute to a diverse dataset.

This data heterogeneity is a challenge for the LPI. The index should handle zeros effectively, as their presence could reflect a true population loss or just temporary detection problems, especially when a single-species population is zero [3,8]. It also needs to be resilient against varying levels of available data, ensuring that missing data, repetitive zeros, and irregular or sparse sampling intervals do not create a trend that does not reflect the state of global biodiversity [2]. Therefore, it is essential for robust interpretation of LPI trends to identify and comprehend how these sources of variation could affect it.

Opinions vary regarding how the LPI responds to these variations. McRae et al.[3] argue that using a comprehensive global dataset of species population time series reduces bias caused by missing data. However, Toszogyova et al. and Troudet et al [7,9] show that uneven availability of population data tends to focus on specific taxa and regions, resulting in substantial taxonomic and spatial biases. Conversely, McRae et al.[10] suggest that applying weights that balance groups (e.g., birds versus reptiles) and regions (e.g., North America versus Africa), across various system configurations, helps provide a more uniform representation of global biodiversity, and Currie et al.[11] contend that explicit testing of weighting and other treatments are critical for the transparency of the methods used in the LPI to monitor long-term biodiversity trends. Nevertheless, Buschke et al.[12] show that the index is highly sensitive to fluctuations in population data. They demonstrate that even with stable values on average, the LPI can show a significant decline due to its calculation method. This indicates a potential bias toward reflecting losses, which may lead to an overestimation of biodiversity loss.

The potential impacts of the data heterogeneity on the robustness of the LPI have been evaluated through the exploration of different methodological choices for index calculation. Currie et al.[11] used the Canadian population data to evaluate how different treatments of zeros, changes in confidence intervals and the uncertainty of the index, the effect of the time series length and the number of data points required, different weighting, and the impact of the baseline on the results. Dove et al.[13] developed a model of trend reliability using simulated datasets as stand-ins for indicator datasets and a distance measure to quantify reliability by comparing partially sampled to fully sampled trends. Marconi et al.[14] used Canadian population data to assess the taxonomic and geographic coverage, as well as the quality of the data in the index. Collen et al.[2] investigated how the use of generalized additive models provides information that enables us to evaluate potential bias in the existing data and to explore the feasibility of data disaggregation based on taxonomic and geographic variations. However, without a baseline that includes complete observations, an absence of zero values, and regular, dense sampling intervals, the effects of these evaluations can't be compared to determine which approach offers the best way to reduce the potential bias from data heterogeneity.

Simulated data provide an opportunity to create a dataset that can be compared with the heterogeneity of the real data. The use of synthetic data provides a high level of control, offering the best-case scenario to assess whether correct estimates in an ecological system can be inferred from real data[15]. It also provides the opportunity to evaluate and compare different properties of our estimators and to evaluate their behaviour and performance[16,17]. In this way, the simulation of stable populations and combining them with empirical data allows us to estimate contributions of missing data or zeros in the LPI result. Here, we use simulated populations for which we can establish the actual trend to jointly evaluate the impact of method and data structure on the sensitivity and robustness of the LPI. We examine whether data removal, presence of zeros, and trend variation influence the simulated LPI trend. Additionally, we analyze the public *LPD* to see how its current structure of missing data and zeros affects the simulated LPI. Our goal is to assess how the method and data structure react to these disturbances.

2. Results

The LPI detects biodiversity trends on stable simulated data.

When the LPI method is applied to simulated population data, it captures population trend. A similar pattern was also observed despite the intentionally introduced variations. The index starts at 1 in the first year and then fluctuates, either increasing or decreasing relative to this baseline [10]. Using a *complete simulated population dataset*, the values across time were close to one (1.0), a number which shows little variation (Fig 1A) and then stability in the populations used. Also, the minimum and maximum coefficient variations are almost identical to the index values, generating a flat line. (Fig 1A). Similarly, inclusion of the three decreasing trend types (concave, convex and linear) produced a consistent trend line that reflected the three types of trends included, following the pattern of the input data without artificial fluctuations (Fig 1B-D). Next, we found that the absence of data (i.e., missing values in the temporal sequences) increased trend variability and could modify the magnitude, but did not alter the direction of the overall trend.

For all the trends analyzed, removing the data produced the expected trend, although there was a noticeable increase in variability due to the gaps created by the missing data (Fig. 1E–G). The 20%, 40%, 60%, and 80% removal scenarios capture the same population trend, highlighting the robustness to data removal, as only the most extreme 95% removal scenario showed a distinct trend (Fig. 1E–G). The greatest variability was seen in the convex trend (Fig. 1F), where the trend showed some increases and decreases over different years, but overall, during the analyzed period, the tendency remained convex. By contrast, the linear decreasing trend showed a consistently linear decline (Fig. 1F), and the concave trend exhibited only minor changes (Fig. 1G).

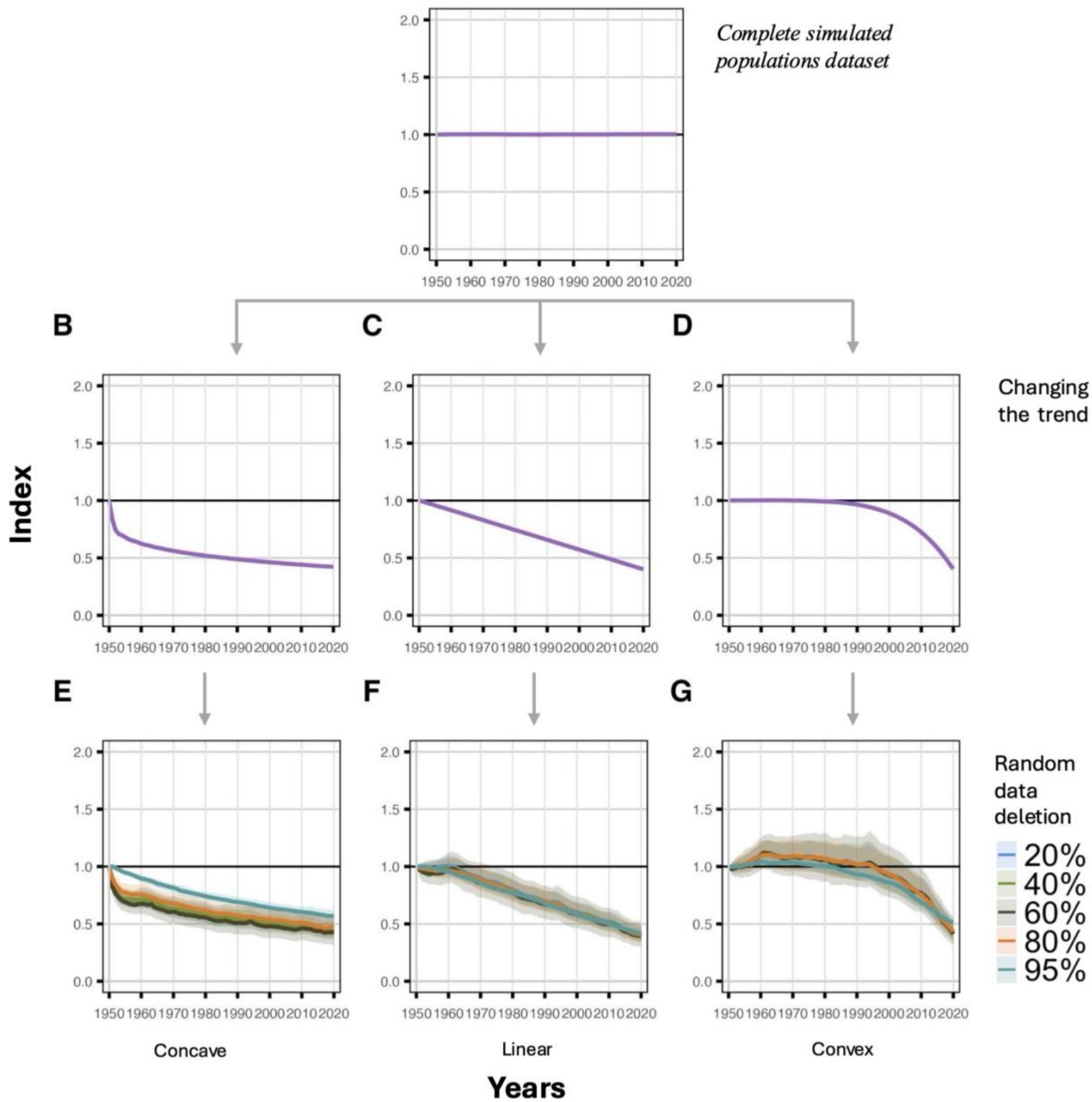

Figure 1. Resulting population trends using simulations for different scenarios used in the study. A) Results obtained by applying the LPI to the *simulated populations dataset* while accounting for sampling error; Results obtained after applying B) concave, C) linear, and D) convex trends using

the full simulated dataset; and E–G) Median values obtained removing randomly 20%, 40%, 60%, 80%, and 95% of the simulated data under E) a concave, F) a linear, and G) a convex trend after 50 iteractions. The values used in the E, F and G obey the median obtained after all of the iterations.

The Living Planex Index shows a consistent decrease when using the Living Planet Database.

Calculating the LPI using the *LPD* without applying diversity weighting (all population time series contributing equally to the index, regardless of taxonomic, realm or geographic region) and representing all populations as distinct species —to ensure that no populations are aggregated or averaged into a single species trend— projects a decrease in population abundances. The index shows fluctuations that dipped and rose by very small amounts, remaining around 1.0, rising slightly between 1960 and 1985 (Fig 2A). Afterward, the index experiences a steady decline. The coefficient of variation is unstable before 1960, increasing between 1950 and 1962, but then remains stable afterward (Fig 2A). These results were verified both by retaining the duplicate records and by removing them, and the results remained the same, which demonstrates the robustness of the workflow used.

The Living Planex Index produces the expected trend (concave, linear and convex) despite missing data, but is affected by both the presence and position of zeros.

We observed that the addition of zeros in the *complete simulated population dataset* distorts the trend generated by the LPI, but that the trends are unaffected by missing data alone (Fig. 2). When we include in the *complete simulated population dataset* both the same zeros and missing data present in the *LPD* (Fig. 2B), the resulting trend was different from the trend observed with our stable simulations (Fig. 1A) and from the empirical data (Fig. 2A). The trend obtained (Fig. 2B) showed an increase from 1960, reached its highest point in 1984, then declined until 2016, when it began to rise again. Next, attempting to separate the effects of zeros and missing data, we observed that removing the data at the same position where it is missing in the *LPD* yielded results identical to the trend observed when we used the *complete simulated population dataset* (Fig 1A and Fig 2D). However, only adding zeros to the dataset at the same positions as the *LPD* caused a decline in the trend. Under this scenario, the trend initially matched that of the completed simulated dataset (Fig. 2C) but diverged later in the series. After 1980, the trend began to drop; there was a more noticeable decline in 1990, it remained relatively stable in 2000, then sharply fell in 2005. Interestingly, after 2011, the trend reversed and started increasing, peaking in 2020 at a level similar to 2005 (Fig. 2C). Finally, when we relocated the zeros in the *empirical population data* with the same annual proportions there are in the *LPD* but across 300 different permutations (while maintaining missing data at the same position), the resulting trend became unpredictable, and the variation was not constant (Fig. 2E).

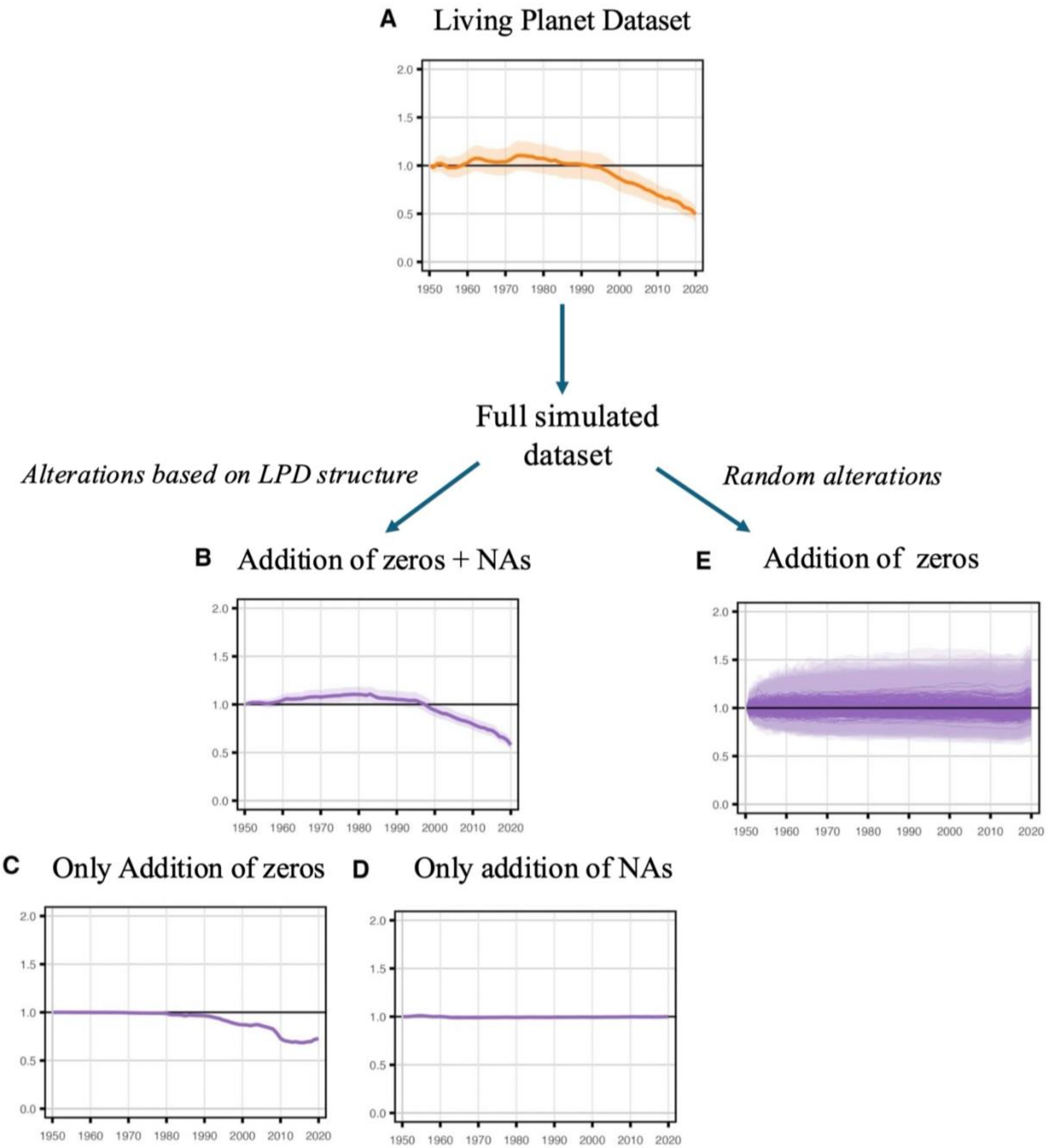

Figure 2. Implications of missing data and the presence of zeros in the *LPD* were assessed using a *complete simulated population dataset* with four approaches. A) Result to calculate the LPI using the *LPD* without including weights and representing all populations as distinct species. B) C*omplete simulated dataset* with data removal at the same locations as missing values in the LPD,

and including zeros at the same positions where they appear sin the *LPD* (see Methods for more information). C) Zeros were added to the *complete simulated population dataset* at the same positions where they appear in the *LPD*. D) Data were removed from the full simulated dataset at the same positions that are missing in the *LPD*. E) The same number of zeros was added to the *LPD*, but they were randomly distributed across 300 different permutations. Note that panels B and C also have confidence intervals, as in the other figures, but all simulation results span a narrower range, making the intervals less visible.

The presence of zeros is a crucial structural feature of long-term time series.

To further explore whether the frequency of zeros influenced index behaviour, we examined the relationship between LPI values and the number of zeros (Fig. 3). The LPI remained close to the reference level (1.0) when the number of zeros was low, but declined progressively as zeros became more frequent. Once the number of zeros exceeded approximately 500, the LPI consistently fell below the reference level. The decline became more pronounced between 1,000 and 1,500 zeros, after which LPI values approached a lower plateau despite continued increases in the number of zeros. At the maximum observed value of 3,401 zeros, the LPI reached its lowest values, although some variability remained (Fig. 3).

The temporal distribution of zeros coincides with variation in the index and, together with our simulation results, suggests that zeros may influence LPI trajectories (Fig. 3 and S1). Early years, particularly before 1997, were characterized by relatively few zeros, whereas subsequent years exhibited a progressive increase. The LPI reached its maximum value in 1974 and then declined gradually, remaining close to 1.0 until the early 1990s despite a steady increase in the number of zeros. By 1997, the cumulative number of zeros had surpassed 1,000, corresponding to an LPI of approximately 0.9. This value doubled to over 2,000 zeros by 2009, while the LPI declined by about 0.2 units. The number of zeros continued to increase rapidly during the following years, reaching a peak of 3,401 in 2013. Although fluctuations occurred after 2010, the overall pattern indicates a persistent decline in LPI associated with an increasing prevalence of zeros (Figs. 3 and S1).

We compared the relationship between the LPI values and the number of zeros across our 300 permutations. The results showed that observed values were generally similar to the simulations when the number of zeros was relatively low. Differences between the *LPD* and simulation results became more apparent at higher numbers of recorded zeros. However, once the count surpassed 1,500 zeros, the results differed between the *LPD* data source and the simulations (Fig. S2).

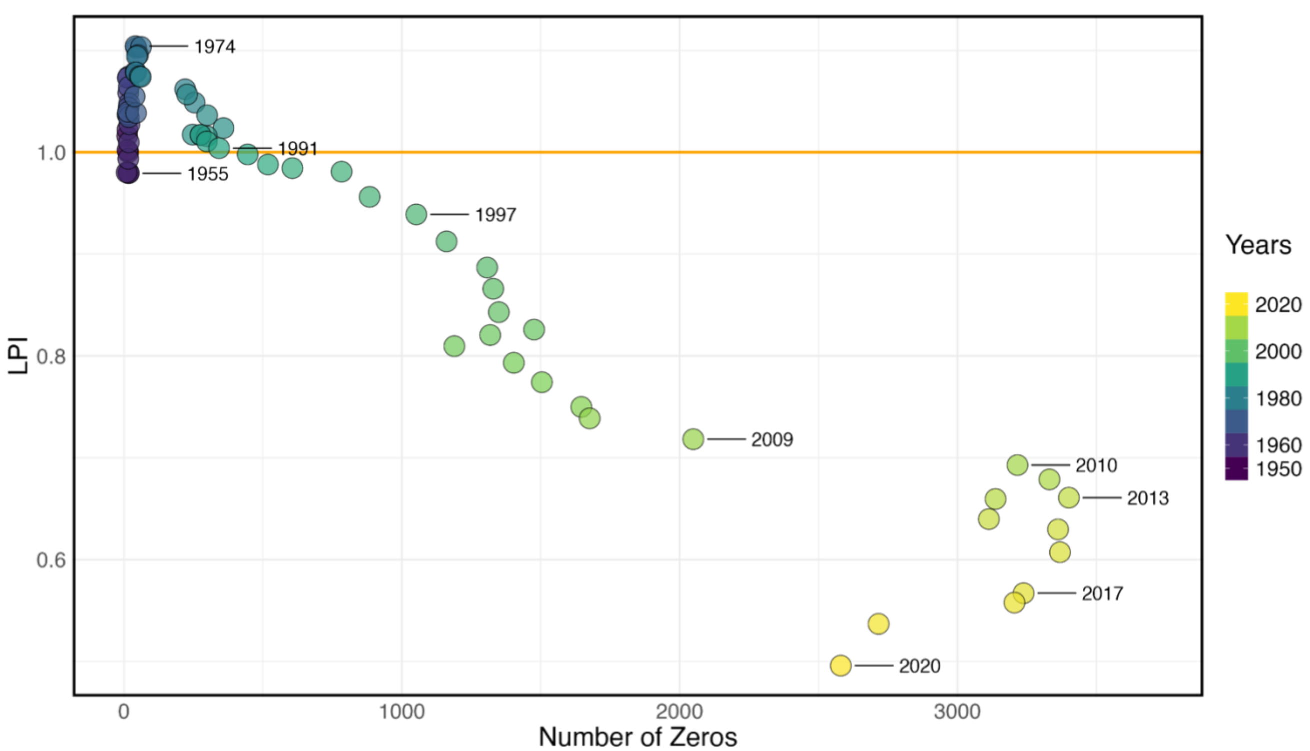

Figure 3. Association between the number of zeros and LPI values between 1950 and 2020 using the *LPD*. The frequency of recorded zeros increased over time and temporally coincided with periods of declining LPI values. Annual changes in the number of zeros were relatively consistent

before 2008, whereas greater year-to-year variation was observed thereafter. Annotated years are shown for narrative purposes to highlight the main temporal trend.

3. Discussion

Using a *complete simulated population dataset*, representing different growth rates and a stable index trend over time, this study demonstrates that the LPI method generally preserves the underlying tendency despite missing data, although it increases its variability. Only when the proportion of missing data exceeded 95% (an extreme scenario) did the removal of data lead to a different trend, underestimating the declining trend from the data. However, our analysis indicates that zeros in the *Living Planet Database* (LPD) —representing real absences in some populations— impact the results and influence trend outcomes based on where they occur. Notably, including zeros from where they are found into the *LPD* in the simulated (stable) populations resulted in discrepancies from expected trends and was sufficient to create decreasing trends in an otherwise stable population due to their temporal position within population time series. Meanwhile, missing data on the *LPD* caused only minor variations when it was included in the *complete simulated population dataset*. Overall, these findings underscore both the importance of testing indicators with a baseline that can help identify limitations in the data used for their calculation. Our results also call for increased effort in identifying the cause and temporal position of zero values in time series that support the LPI calculation.

Using a *complete simulated population dataset* as a reference, our study reveals that the LPI method is highly sensitive to the number and temporal positions of zeros present in the *LPD* across all populations considered. These zeros —when occurring in their observed temporal positions— have a substantial impact, both qualitative and quantitative, on the final trend, and are sufficient to explain the tendency observed in the simulated LPI (shown in Fig. 2A). Similar to Toszogyova et al.[7], our analyses show that zeros can lead to significant drops or spikes in the index, when they are temporally structured within population time series. Three factors likely explain this outcome: firstly, the geometric mean, the method used to aggregate the LPI and similar multi-species indices, is sensitive to the presence of zero values, which presents a challenge to capturing change in rare species or those hard to detect.[8] Secondly, the zero values present in the *LPD* are more prevalent in the second half of the time-series, with the increase in the number of recorded zeros temporally coinciding with periods of declining simulated LPI values (Fig. 2B-C and Fig. 3).

To address effect of zeros in a multi-species index, several alternatives have been proposed in other papers, including treating zeros as missing (i.e. NA) [7,11,14], adding a small constant such as 1% of the mean [2,18], the minimum value, the number 1, or an extremely small value (e.g., 0.000001), as well as conditional replacements where leading, middle, and trailing zeros are substituted with combinations of NA and 1% of the mean [11,2]. In one study, these modifications result in different trends, causing a variation in the outcome as well as in the confidence intervals.[11]

This issue is particularly challenging because, in ecological datasets, the presence of zeros can originate from different processes. Zeros may represent true biological absences, but they may also arise from imperfect detectability or sampling limitations, making it difficult to distinguish ecological signal from observational noise. Multiple trailing zeros are more likely to represent a local extinction, an important signal for a biodiversity indicator. Therefore, correctly classifying zeros is crucial to avoid false zeros that could hinder real patterns or lead to misleading conclusions about population change, a concern that has been analyzed in the literature[19]. Wenger & Freeman (2008), through practical examples, have investigated the explicit effect of zeros and demonstrated how the adoption of methods that explicitly model the sources of zero observations improves inference and strengthens robustness of ecological analyses. The *LPD* and resulting LPI, however, does not currently incorporate this approach. Instead, the zeros are transformed by adding 1% of the mean to all values in the time series, in order to calculate rates of change for the index on a log scale[2]. As a result, the influence of zero values on index trends appears to be driven more by the properties of the transformation than by the data type itself, an important consideration when interpreting apparent declines in the LPI.

Similar to zero values, missing values have been identified as a potential source of bias in the LPI due to the way they are incorporated into index calculations[7]. In practice, the LPI accounts for missing values through methodological transformations designed to standardize trajectories over time. Therefore, if a species is not observed in a particular year, values are interpolated using different modelling approaches but only within the bounds of the first and last year of the time-series [2]. Our analysis shows that removing some of the simulated data causes variations (Fig. 1D-E), but the overall trend stays consistent with the baseline.

The only exception was when we removed 95% of the data, at which point the LPI maintained but started to underestimate the declining trend. Additionally, missing data in the *LPD* does not alter the overall trend shape once added to the simulated stable populations (Fig. 2), indicating a limited impact on the overall outcomes alone. However, our results suggest that missing data do alter the impact that zeros have on the index (Fig. 2B), as in our example, the index trend shifts to a more positive, then a more negative trend compared to the stable baseline. Previous works have shown that random fluctuations naturally distort LPI trends, often biasing toward decline, and should be evaluated against null-model baselines [12]. Although these random fluctuations were not explicitly included in our analysis, the persistence of minor variations when we analyze the effect of missing data supports that the index can maintain the trend despite missing data.

Addressing these challenges arising from zeros and missing data requires models with clearly defined baselines to evaluate how the index responds to heterogeneous data and to ensure that observed patterns reflect genuine population trends rather than methodological artifacts. Zeros play a critical role in calculating the Living Planet Index (LPI) because their position can substantially influence estimated trends, even when population time series are complete and show no apparent declines. It is therefore essential to determine whether zeros represent true observations or arise from missed observations before including them in the index. Although the LPI can be informative when data are sparse, limited or discontinuous time series reduce accuracy and increase trend variability, highlighting the need for more continuous monitoring to improve robustness. Validation using both simulated and well-characterized ecological datasets is particularly important, as it can help identify sources of variation, reduce arbitrary treatments of zeros, and reveal data gaps that affect trend detection. While both the observed temporal trend and the final value of the index may reflect the influence of zeros or near-zero values, the LPI still indicates a downward trajectory after zeros are removed and replaced by missing data. We have demonstrated the value of using simulated data to better understand how the LPI responds to zero values and missing data, both of which are pervasive features of population timeseries data, and more likely to increase as the timespan and the taxonomic diversity of the dataset to which the LPI is applied increase.

Long-term research is a cornerstone of ecological studies[20,21], as it can capture long-term changes in biodiversity that sporadic collections of samples cannot. Nevertheless, our comparisons indicate that increasing numbers of recorded zeros may influence LPI trajectories (Fig. 3). As long-term datasets expand, understanding whether zeros represent true absences, local extinctions, non-detections, or other observation processes becomes increasingly important for interpreting biodiversity trends. This suggests a need for improving long-term research that captures and analyzes data[22], primarily by more adequately documenting the process that leads to a population size being registered as a zero. Introducing regular monitoring programs using structured protocols can help mitigate gaps in our data streams, ensuring a more comprehensive understanding of the systems we're monitoring[21]. By adopting a long-term perspective that balances the need for frequent updates with the importance of uninterrupted data collection, we can better appreciate the significance of zeros as a structural feature of time series data.

We recommend determining the nature of the zero values in the empirical LPI to inform appropriate treatments. Because the trend reported by the LPI responds, in addition to genuine changes in species abundance, to the temporal position of zeros in the time series, we encourage users of the method to carefully distinguish between actual zeros (the population is reliably absent/locally extinct at this time point) from artifactual zeros (used in place of missing data). Further research is needed to assess how the diversity of species or populations included affects index performance, particularly to determine minimum data requirements across spatial scales, biogeographic realms, time series coverage, and taxonomic groups.

5. Methods

We used a population-simulation matrix to assess the Living Planet Index's sensitivity to the presence and position of zeros and to missing data (NA values). The matrix represents a dataset showing the hypothetical number of individuals in a not-real population across different generations. For methodological purposes, we consider each generation as 1 year. Then, we applied controlled perturbations to the data quality to see if the LPI maintained the initial trend. Deviations in indicator outcomes from the population simulations were used to assess robustness. The complete workflow can be observed in Figure 4.

Living Planet Index calculation using the Living Planet Database

We used the latest available public version of the *Living Planet Database* (*LPD*)[23], the largest repository of vertebrate population abundance data, to create a trend based on genuine data. This public version contains data from 35,996 populations of 5,177 species, sourced from scientific papers, online databases, government, and expert reports[24]. As our aim was to evaluate the LPI using baseline simulated data, we included the full number of years reported in the *LPD*. We followed the published LPI method for calculating an index [2]. In order to compare with simulated data, we represent all populations as distinct species to ensure that no populations are aggregated or averaged into a single species. Also, as our goal wasn't to study the impact of data availability, we didn't include the weighting that is used for the global LPI, which assigns weights to populations based on the regions, realms, and taxa where they were monitored (McRae et al., 2017).

Simulation overview

We used simulated population data to assess the sensitivity of the Living Planet Index (LPI) method and data structure. The matrix shows a *complete simulated population dataset* which reflects a stable population, indicating that the simulated populations remain consistent across generations. Using the *complete simulated population dataset*, we applied controlled perturbations to the data quality to see if the LPI maintained a similar trend. Deviations in indicator outcomes from the population simulation were used to assess robustness. The complete workflow can be observed in Figure 4.

Population Simulated Data: To simulate each population, we used the discrete logistic growth equation proposed by Richel in 1954 (Equation 1). It models density-dependent regulation in populations with discrete generations by linking the population size at one time step to the next, based on the intrinsic growth rate and carrying capacity [25].

$$N_{t+1} = N_t \cdot e^{r \cdot (1 - \frac{N_t}{K})} \quad (1)$$

Where N_t represents the population size at time t; r the intrinsic rate of natural increase; K is the carrying capacity, representing the maximum sustainable population size the environment can support. Finally, e is the base of the natural logarithm (≈ 2.718). Simulations were run on 50,000 independent hypothetical populations over 500 generations.

The initial population sizes simulated were randomly selected, ranging from a dozen to thousands, representing various population sizes. The number of generations was chosen to ensure enough time for populations to reach equilibrium or show stochastic fluctuations. The intrinsic growth rate values ranged (r) from 0.05 to 0.45, reflecting the minimum and maximum growth rates documented in the literature for vertebrates[26,27]. To set the carrying capacity (K), we used the minimum value based on the Elk populations (~<10 per herd in equilibrium)[28], and the maximum estimate derived from Salmon (~71,844 per delta)[29]. To account for environmental differences among populations, these parameters were varied across replicate populations, allowing each to have its own distinct growth rates and carrying capacities[30,31]. To obtain a *complete simulated population dataset* comparable with the *empirical population data*, we randomly selected the number of populations and generations of our simulated dataset, matching the *LPD* dataset's structure in terms of population units (rows) and years (columns).

Scenarios

Observed error: To evaluate the impact of measurement uncertainty on observed population variability, an extra layer of randomness was added to the simulated data to simulate sampling error[32]. This was achieved by randomly modifying population values to account for detection or counting errors, with a 0.1 probability to introduce more sampling variation. This was done by randomly adjusting population values to emulate inaccuracies or imperfect detection in our simulated population data. Incorporating observation error allows evaluation of how sampling variability can alter perceived population trends and indices, even when underlying dynamics remain stable.

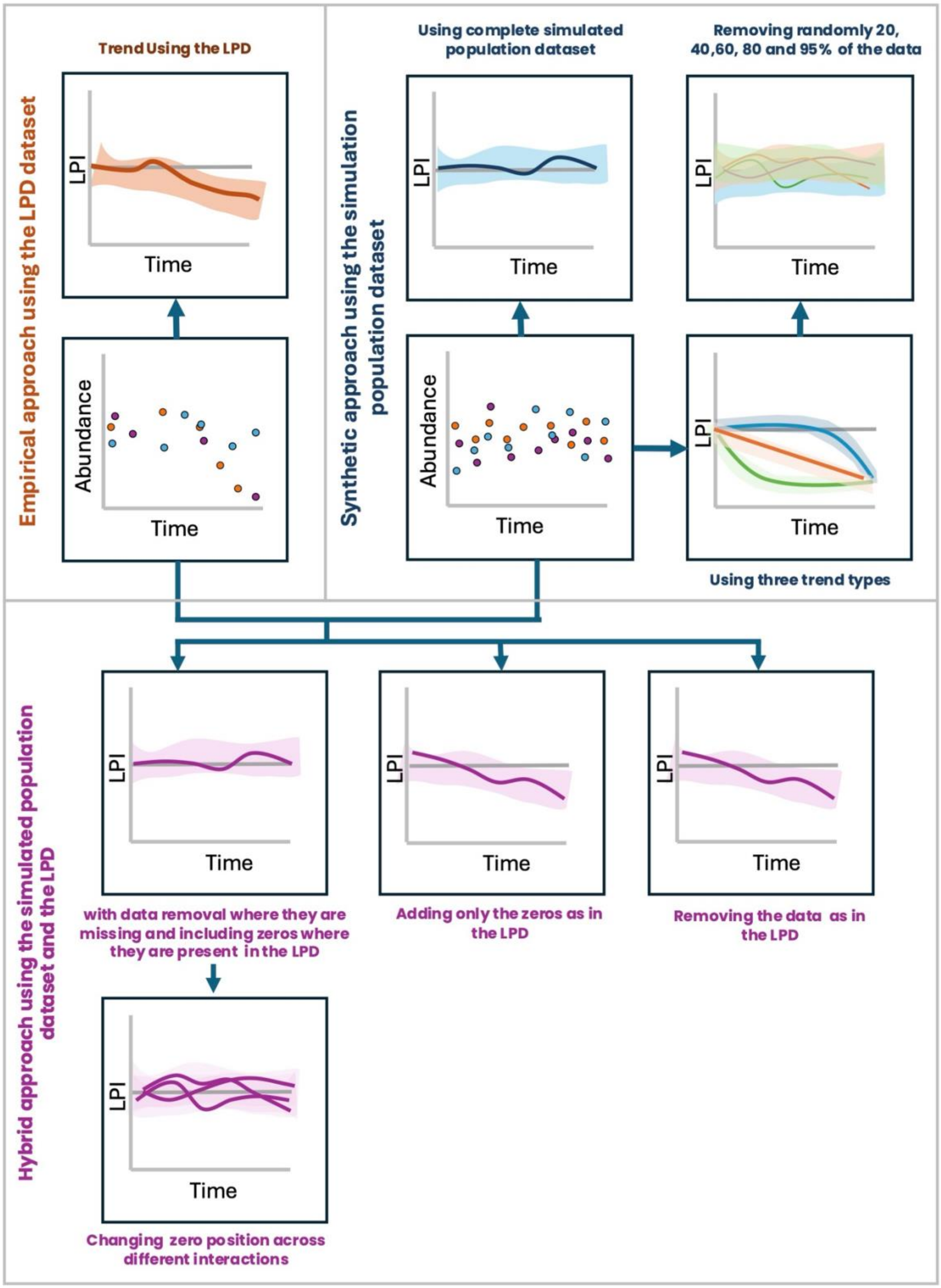
Empirical approach using the LPD dataset
Trend Using the LPD
LPI
Time
Abundance
Time
Synthetic approach using the simulation population dataset
Using complete simulated population dataset
LPI
Time
Abundance
Time
Removing randomly 20, 40,60, 80 and 95% of the data
LPI
Time
LPI
Time
Using three trend types
Hybrid approach using the simulated population dataset and the LPD
LPI
Time
with data removal where they are missing and including zeros where they are present in the LPD
LPI
Time
Adding only the zeros as in the LPD
LPI
Time
Removing the data as in the LPD
LPI
Time
Changing zero position across different interactions

Figure 4. Workflow of the approach used in this study. The empirical approach refers to the calculation of the trend using the data available at the *LPD*. The synthetic approach refers to the trends generated using the *complete simulated population dataset*. The results include the types of trends and also remove 20%, 40%, 60%, 80%, and 95% of the dataset at random. Finally, the hybrid approach represents the use of the *complete simulated population dataset*, but with data removed as missing according to the structure in the LPD, adding zeros and changing their position across interactions per row.

Modifying the trend: We generated three hypothetical population decline trajectories —concave, linear, and convex— and subsequently applied them to the *complete simulated population dataset* to reflect the diversity of trend types observed in nature (fig. 4). These trajectories were used because even random fluctuations and non-linear trends (such as concave or convex shapes) can influence the behaviour of overall biodiversity indices[12]. This approach allowed us to test if changes in population dynamics affect the stability and interpretation of the simulated results.

Missing data: Missing observations (NA values) in long-term population data can affect trend detection and create bias interference[33], as well as reduce the statistical power and hamper inference (Łopucki et al., 2022). For that reason, we randomly removed 20%, 40%, 60%, 80%, and 95% of the data points for the whole dataset (fig. 4). This allowed us to examine how data loss might bias the trend captured by the LPI. We conducted these analyses using the same trend as in the last step[12] to assess whether combining different trends with missing data could lead to unforeseen scenarios that the LPI couldn't detect. Finally, we conducted 50 iterations for each combination to evaluate the robustness of the results with respect to the locations of the data points removed. We reported the median, minimum, and maximum index values from all iterations (fig. 4).

Incorporating null and zero values in the simulated data: We utilized the *LDP* structure and incorporated it into our *complete simulated population dataset* to examine whether the structure of the empirical dataset possessed particular features that affect the trend [3,10]. First, we eliminated data points at identical time–series positions that were absent in the empirical dataset per row, ensuring the simulated gaps mirrored the real-world missingness pattern. Second, per row, we replaced values at those same positions with zeros to simulate cases where monitoring records are true zero observations or non-detections (fig. 4).

After we evaluated how the distribution of missing or zero values could affect trend detection, we performed 300 permutations as follows. First, we replace the zeros and remove the data from the same cells that are in the *LPD* in the iterations in the *complete simulated population dataset*. After, we shuffled the position of the zeros randomly within the time period covered by each time series, but we maintained fixed cells with missing data in the matrix with the *complete simulated population dataset*. With the resulting matrix, we applied the LPI method to calculate an overall trend. We repeated this process 300 times. In each iteration, the zeros were shuffled per row.

All data analysis was performed in R (version 4.3[34]) using the main packages: data.table (v1.16.2[35]), future.apply (v1.11.3[36]), and future (v1.40.0[36]) for efficient data handling and parallel computation; ggplot2 (v3.5.2[37]), patchwork (v1.2.0[38]), and RColorBrewer (v1.1-3[39]) for data visualization; and missMethods (v0.4.0[40]) and tidyverse (v2.0.0[41]) for data preprocessing. The Living Planet Index was calculated using the rlpi package (v0.1.0[42]).

5. Data availability

The study relies on the *Living Planet Database (LPD)*, which is publicly available through the official Living Planet Index data portal: https://www.livingplanetindex.org/data_portal. All datasets used in the analyses can be accessed and downloaded from this resource.

6. Code availability

All code used for the population simulations and the analytical routines described in this study is openly available on GitHub at the following repository: https://github.com/crcruzr/LPI_Simulations/. The repository includes scripts, documentation, and instructions to reproduce the results.

7. Acknowledgments

This manuscript results from a working group supported by the Living Data Project (LDP), an initiative of the Canadian Institute of Ecology and Evolution funded through a Collaborative Research and Training Experience (CREATE) grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). We also thank the reviewers for their valuable feedback during the development of this document. We acknowledge that this study was partially conducted on land within the traditional unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat and Omàmiwininiwak nations.

8. Author contributions

All authors contributed substantially to manuscript revision, provided critical feedback, and approved the final submitted version.

9. Competing interests

The authors' institutions did not affect the study design, data analysis, interpretation of results, or the decision to publish. All authors declare no competing interests.

10. References

1. Ledger, S. E. H. *et al.* Past, present, and future of the Living Planet Index. *Npj Biodivers.* **2**, 12 (2023).
2. Collen, B. *et al.* Monitoring change in vertebrate abundance: the living planet index. *Conserv. Biol. J. Soc. Conserv. Biol.* **23**, 317–327 (2009).
3. McRae, L. *et al.* The utility of the Living Planet Index as a policy tool and for measuring nature recovery. *Philos. Trans. R. Soc. B Biol. Sci.* **380**, 20230207 (2025).
4. McRae, L. *et al. Living Planet Index Technical Supplement 2024*. https://osf.io/6te9h/ (2024).
5. Deinet, S., Marconi, V., Freeman, R., Puleston, H. & McRae, L. Living Planet Index technical supplement 2024. https://doi.org/https://www.google.com/url?q=https://doi.org/10.17605/OSF.IO/6TE9H&sa=D&source=docs&ust=1775937926621981&usg=AOvVaw0r8YXVJTCqeMkw73kRj_JN (2024) doi:https://www.google.com/url?q=https://doi.org/10.17605/OSF.IO/6TE9H&sa=D&source=docs&ust=1775937926621981&usg=AOvVaw0r8YXVJTCqeMkw73kRj_JN.
6. Moussy, C. *et al.* A quantitative global review of species population monitoring. *Conserv. Biol.* **36**, e13721 (2022).
7. Toszogyova, A., Smyčka, J. & Storch, D. Mathematical biases in the calculation of the Living Planet Index lead to overestimation of vertebrate population decline. *Nat. Commun.* **15**, 5295 (2024).

8. Korner-Nievergelt, F. *et al.* Multi-species population indices for sets of species including rare, disappearing or newly occurring species. *Ecol. Indic.* **140**, 109005 (2022).
9. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* **7**, 9132 (2017).
10. 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).
11. Currie, J. *et al.* Navigating methodological decisions: balancing rigor and data volume of the Canadian Living Planet Index. *Facets* **11**, 1–14 (2026).
12. 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).
13. Dove, S., Böhm, M., Freeman, R., McRae, L. & Murrell, D. Quantifying reliability and data deficiency in global vertebrate population trends using the Living Planet Index. *Glob. Change Biol.* https://doi.org/10.1111/gcb.16841 (2023) doi:10.1111/gcb.16841.
14. Marconi, V. *et al.* Population declines among Canadian vertebrates: But data of different quality show diverging trends. *Ecol. Indic.* **130**, 108022 (2021).
15. Zurell, D. *et al.* The virtual ecologist approach: simulating data and observers. *Oikos* **119**, 622–635 (2010).
16. DiRenzo, G. V., Hanks, E. & Miller, D. A. W. A practical guide to understanding and validating complex models using data simulations. *Methods Ecol. Evol.* **14**, 203–217 (2023).
17. Lotterhos, K. E., Fitzpatrick, M. C. & Blackmon, H. Simulation Tests of Methods in Evolution, Ecology, and Systematics: Pitfalls, Progress, and Principles. *Annu. Rev. Ecol. Evol. Syst.* **53**, 113–136 (2022).

18. Burns, F. *et al.* An assessment of the state of nature in the United Kingdom: A review of findings, methods and impact. *Ecol. Indic.* **94**, 226–236 (2018).
19. Martin, T. G. *et al.* Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecol. Lett.* **8**, 1235–1246 (2005).
20. Kuebbing, S. E. *et al.* Long-term research in ecology and evolution: a survey of challenges and opportunities. *Ecol. Monogr.* **88**, 245–258 (2018).
21. Lindenmayer, D. B. *et al.* Value of long-term ecological studies. *Austral Ecol.* **37**, 745–757 (2012).
22. Festa-Bianchet, M., Douhard, M., Gaillard, J.-M. & Pelletier, F. Successes and challenges of long-term field studies of marked ungulates. *J. Mammal.* **98**, 612–620 (2017).
23. LPI. Living Planet Index database. (2024).
24. Ledger, S. E. H. Past, present, and future of the Living Planet Index. https://doi.org/10.1038/s44185-023-00017-3 (2023) doi:10.1038/s44185-023-00017-3.
25. Ricker, W. Stock and Recruitment. *J. Fish. Res. Board Can.* **11**, 559–623 (1954).
26. Pardo, S. A. & Dulvy, N. K. Body mass, temperature, and depth shape the maximum intrinsic rate of population increase in sharks and rays. *Ecol. Evol.* **12**, e9441 (2022).
27. Dillingham, P. W. *et al.* Improved estimation of intrinsic growth rmax for long-lived species: integrating matrix models and allometry. *Ecol. Appl.* **26**, 322–333 (2016).
28. Koetke, L. J., Duarte, A. & Weckerly, F. W. Elk population dynamics when carrying capacities vary within and among herds. *Sci. Rep.* **10**, 15956 (2020).
29. Hall, J. *et al.* Estimating Juvenile Salmon Estuarine Carrying Capacities to Support Restoration Planning and Evaluation. *Estuaries Coasts* **46**, 1046–1066 (2023).

30. Dennis, B. & Ponciano, J. M. Density-dependent state-space model for population-abundance data with unequal time intervals. *Ecology* **95**, 2069–2076 (2014).

31. Gotelli, N. J. & Gotelli, N. J. *A Primer of Ecology*. (Oxford University Press, Oxford, New York, 2008).

32. Dennis, B., Ponciano, J. M., Lele, S. R., Taper, M. L. & Staples, D. F. Estimating Density Dependence, Process Noise, and Observation Error. *Ecol. Monogr.* **76**, 323–341 (2006).

33. Onkelinx, T., Devos, K. & Quataert, P. Working with population totals in the presence of missing data comparing imputation methods in terms of bias and precision. *J. Ornithol.* **158**, 603–615 (2017).

34. R Core Team. R: A Language and Environment for Statistical Computing. https://www.R-project.org/ (2023).

35. Barrett, T. *et al.* data.table: Extension of `data.frame`. https://CRAN.R-project.org/package=data.table (2024).

36. Bengtsson, H. A Unifying Framework for Parallel and Distributed Processing in R using Futures. *R J.* **13**, 208–227 (2021).

37. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

38. Pedersen, T. patchwork: The Composer of Plots. https://CRAN.R-project.org/package=patchwork (2024).

39. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. https://CRAN.R-project.org/package=RColorBrewer (2022).

40. Rockel, T. missMethods: Methods for Missing Data. https://CRAN.R-project.org/package=missMethods (2022).

41. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
42. Freeman, R., McRae, L., Deinet, S., Amin, R. & Collen, B. rlpi: Tools for calculating indices using the Living Planet Index method. https://github.com/Zoological-Society-of-London/living_planet_index (2024).

Supplementary Material

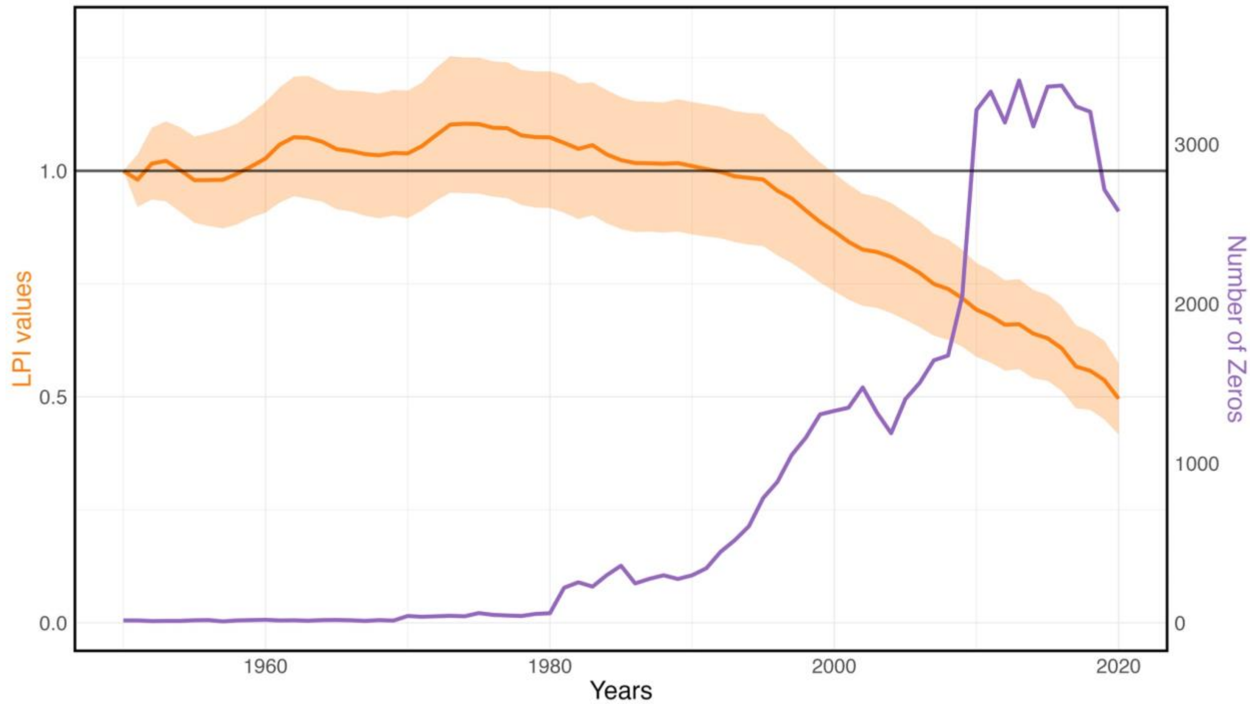

Figure S1. Changes between the LPI (orange) and the number of zeros (purple) per year. The comparison between trajectories presents an inverse relationship between the LPI values and the Number of zeros.

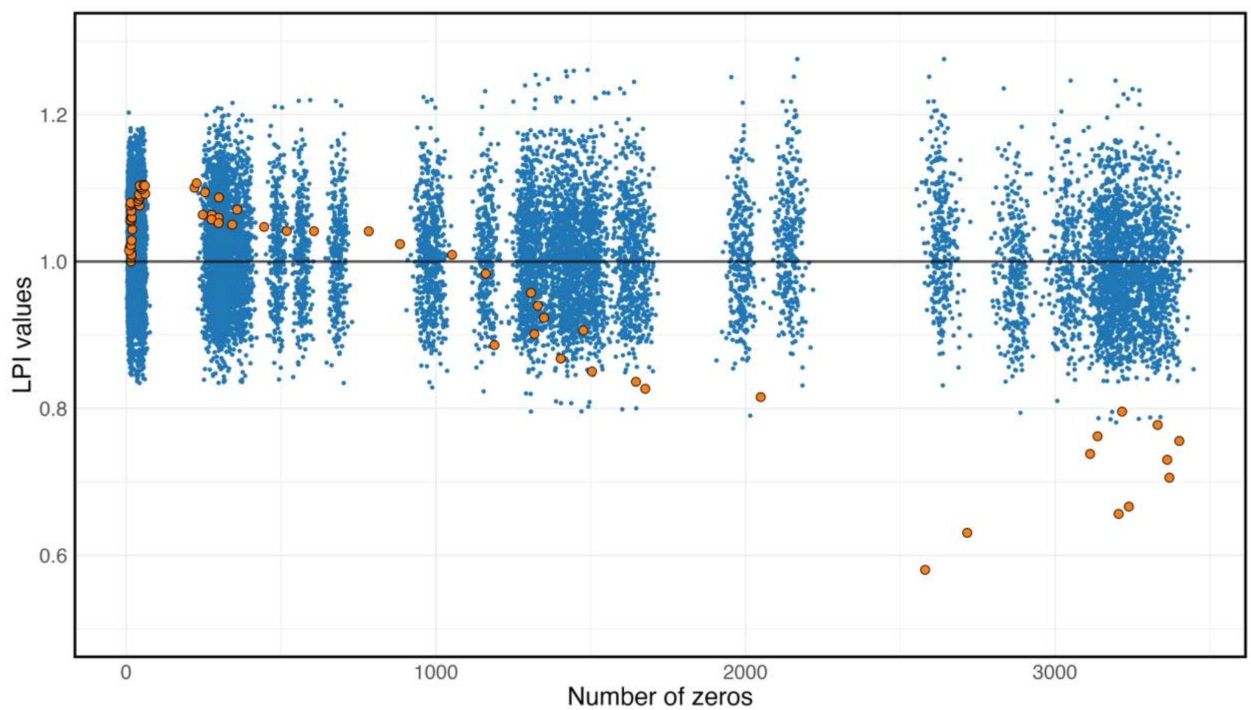

Figure S2: The relationship between the number of zeros and LPI values. Blue illustrates the relation between the two metrics across 300 permutations, while orange dots reflect the relationship using the *LPD*.