Title: How much monitoring is needed to reliably track progress towards genetic diversity targets?

Author names: Katherine Hébert[1,2], Laura Pollock[1,2], Sean Hoban[2,3,4]

Affiliations:

[1] Department of Biology, McGill University, Montreal, Canada.

[2] Quebec Centre for Biodiversity Science, McGill University, Montreal, Canada.

[3] Center for Tree Science, The Morton Arboretum, Lisle, 60532, USA.

[4] Committee on Evolutionary Biology, The University of Chicago, Chicago, 60637, USA.

Keywords: genetic diversity, indicator, Global Biodiversity Framework, performance test, biodiversity change, biodiversity monitoring, conservation

Corresponding author: Katherine Hébert (katherine.hebert@mcgill.ca)

Address: 1205 Dr Penfield Ave, Department of Biology, McGill University, Montreal, Quebec, Canada H3A 1W9

Abstract

Achieving global biodiversity targets hinges on indicators of biodiversity change that convert raw data into reliable numbers that can shape policy, conservation, management, and, ultimately, the future of biodiversity worldwide. Indicators can only be used confidently if they detect and summarise biodiversity trends as intended, given the available data worldwide. Knowing whether indicators can reliably detect and summarize trends as intended requires robust testing, which is a challenging and under-developed practice. Here, we test the performance of a genetic diversity indicator of the Global Biodiversity Framework (GBF), the Proportion of populations with an effective size greater than 500 (or, Ne>500) and show that it can be reliably reported under realistic scenarios of population trends, monitoring frequency, and observer error. To ensure this reliability, our results suggest that monitoring programs aim to monitor populations every 1 to 4 years, at least 40% of populations per species, and at least 8% (species pools of several thousands), 23% (species pools 300 to 500 species) or 56% (species pools <100) of the targeted species richness. These findings show that the indicator is, in addition to being feasible and meaningful, technically reliable under realistic biodiversity monitoring schemes. Going forward, it is still essential to invest in genetic monitoring using indicators and DNA-based data, given the goal of safeguarding genetic diversity in the Global Biodiversity Framework. Beyond this indicator, we emphasize that performance testing is needed for more indicators to ensure reliable progress tracking towards the GBF targets and the global goal of halting biodiversity loss.

1. Introduction

Governments, civil society, and indigenous peoples are working to stem biodiversity loss and restore society's relationship with nature. Building on the lessons learned from prior biodiversity commitments, the Parties to the Convention on Biological Diversity (CBD; 195 countries plus the European Union) committed to a new Global Biodiversity Framework (GBF) in December 2022, in which indicators play a key role to measure progress towards goals and targets (Affinito et al., 2024). Indicators are summaries of changes in the state of a system which are used to inform decision making (Rowland et al., 2021) and to identify and act on the causes of these changes (Jones et al., 2011). They translate raw data about biodiversity (e.g. extinction risk, species protection, genetic diversity, etc.) into simple metrics that are both ecologically and politically meaningful to track progress towards targets and to inform policy decisions (Noss, 1990; Jones et al., 2011). Most indicators under the GBF need further guidance and capacity building according to the CBD's Ad Hoc Technical Expert Group on Indicators (CBD/SBSTTA/26/2, 2024), notably to design monitoring programs that fulfill each indicator's needs (Hoban et al., 2014).

To fill this gap, there have been calls for evaluations to ensure indicators can be reported reliably for decision support and progress tracking (Nicholson et al., 2021; Stevenson et al., 2024; Leung & Gonzalez, 2024; Hébert et al., 2025). An indicator's reliability depends on its accuracy, which is a measure of the departure or deviance of the estimated indicator from the expected "true" value, and bias, which refers to whether the indicator systematically under- or over-estimates the expected "true" value (Hellmann & Fowler, 1999). Reliability also depends on uncertainty in the indicator value, arising from various sources of variation in the data and how data is processed and aggregated into the indicator (Hébert & Gravel, 2023). Measuring uncertainty and comparing it to an "acceptable" level is important to ensure indicators are interpreted with an appropriate degree of confidence when informing decisions, investments, and policies (Fischhoff & Davis, 2014; Leung & Gonzalez, 2024). Because indicators may perform unpredictably, simulations are often needed to assess the potential impacts of biodiversity change and data quality or quantity on the indicator's reliability (Flather et al., 1997; Visconti et al., 2016; Nicholson et al., 2019; Stevenson et al., 2021). Power analyses can help identify the quantity and quality of data that is needed for an acceptable level of uncertainty risk (Leung & Gonzalez, 2024). With some exceptions (Leung & Gonzalez, 2024; Toszogyova et al., 2024), most biodiversity indicators have not yet been evaluated in terms of how reliably they detect and summarise biodiversity changes.

Biodiversity is composed of ecosystems, species, and genes. Genetic diversity is the most fundamental because it is the basis of adaptation and resilience for wild and domesticated species and contributes to mitigating climate change impacts, supporting ecological restoration, and stabilizing food security. Monitoring the genetic diversity of populations is critical, as it can motivate action early in the timeline of biodiversity collapse to prevent further losses at the species, community, or ecosystem levels (Cerini et al., 2023). For these reasons, the Kunming-Montreal Global Biodiversity Framework set objectives for maintaining and restoring the genetic diversity of wild and domesticated populations by 2030 (Target 4) and by 2050 (Goal A), to ensure species' adaptive potential (Robuchon et al., 2023). One of the Headline Indicators (A.4) of the GBF, "Proportion of populations with an effective size (Ne) >500" (referred to hereafter as Ne>500 indicator), tracks genetic diversity to measure progress towards these goals (Hoban et al., 2023, 2024).

Effective size (Ne) is the genetic analogue of a population's census size (Nc): it relates to the genetic behavior of a population and especially the maintenance of the population's adaptive capacity. The basis of the Ne>500 indicator, an effective size of 500, is a recognized threshold for maintaining genetic diversity within populations over the long term ((Franklin & Frankham, 1998).This size helps populations adapt more efficiently to environmental change via natural selection, and with less demographic cost. The threshold of Ne 500 is pragmatic yet imperfect. Other Ne thresholds have been highlighted as important (Lynch & Lande, 1998), and any decrease in Ne causes some genetic erosion (Allendorf et al., 2024). Moreover, any threshold-based indicator results in a binary outcome which has unique properties (a point we will expand on in the Discussion), but this is a common approach in many indicators (e.g. income level for poverty, population size thresholds for the IUCN Red List). Despite these imperfections, Ne 500 is an approximate threshold of accelerating rate of genetic diversity loss, making it a useful heuristic to identify populations in danger of declines and losing their adaptive capacity, pointing to where conservation action is needed (Frankham, 2022; Willi et al., 2022).

Effective size (Ne) is flexible to the type of data available: it can be calculated from DNA-based data if sampling is sufficient and certain assumptions are met (which are not always met, see Santos del Blanco et al. (2022)), or it can be obtained from population census data by multiplying census size (Nc) by an Ne/Nc ratio of approximately 0.1 (Hoban et al., 2021) when genetic data is not available. This flexibility means that the Ne>500 indicator (Headline A.4) is immediately feasible and can incorporate data about more populations and species than it would be possible to include using DNA-based data alone. After four years of development and consultation with CBD stakeholders, the Ne>500 indicator was demonstrated at a large scale, for over 900 species and 5000 populations in nine countries, largely based on population census data

(Hoban et al., 2023; Mastretta-Yanes et al., 2024). Because regularly repeated, multi-population sampling of DNA data is currently and will likely be rare until at least 2030 even in high income countries (Posledovich et al., 2021; Mastretta-Yanes et al., 2024), the indicator will likely continue to be calculated with population census data until genetic monitoring is more common.

We must then understand how the Ne>500 indicator performs when confronted with the most common data it is applied to, which is population census data (Nc), under the assumption that the Ne/Nc ratio is known. More specifically, policymakers have asked: (1) how accurately the indicator represents populations on the ground, (2) how responsive the indicator is to conservation interventions and management choices, and (3) what quantity and quality of data are needed to ensure the indicator's accuracy, knowing that it unrealistic to monitor all populations of a species and all species of a country. Several factors could impact the indicator's accuracy and detection power, including monitoring frequency (years between monitoring events), intensity (how many populations and species are monitored), and duration, as well as observer error (the degree to which observed Nc departs from true Nc due to variation in detection, sampling effort, etc.), population growth rate (a product of the species' biology and the environment, including anthropogenic pressures), and how these factors relate to population size (Yoccoz et al., 2001; Clark & Bjørnstad, 2004; Reed & Hobbs, 2004; Wauchope et al., 2019; White, 2019; Bennett et al., 2024). Identifying how these factors influence the indicator's performance is important to allocate resources to species that need genetics-informed management or protection.

In this paper, we evaluate the accuracy and detection power of Headline A.4 Indicator "Proportion of populations with an effective size >500" (hereafter referred to as "Ne>500 indicator") in simulated scenarios of biodiversity change under different monitoring schemes.

More specifically, we use simulations to assess the indicator's power to detect changes from low (infrequent and inaccurate) to high quality (frequent and accurate) data. We then determine how much data is needed to calculate the indicator for a species and for a country with an acceptable level of accuracy, based on simulated and empirical data. From these evaluations, we make practical recommendations for data selection and monitoring programs to support reliable indicator reporting.

2. Methods

To assess the Ne>500 indicator's ability to detect changes in the data it is typically applied to, we will test the indicator's performance with Ne derived from population census size (Nc) time series, under simulated scenarios of biodiversity change and monitoring frequency, accuracy, and intensity. To evaluate indicator performance, we measured accuracy, i.e. how closely the estimated Ne>500 indicator value approaches the expected "true" value (Hellmann & Fowler, 1999), under different conditions of biodiversity change, observer error, and monitoring frequency. Because the "true" value is inherently impossible to access in empirical data, we use simulations to generate the "true" and "observed" values of a population's trend (See Fig. S1 for simulated population trends).

To keep our findings relevant to monitoring under the Convention on Biological Diversity, European Habitats Directive, and other national and international biodiversity commitments, we prioritized realistic parameters for the population simulations that are grounded in empirical data or that are relevant for conservation interventions aiming to halt and reverse genetic diversity loss (Supplementary Material S1). To focus on the effects of census data quality and quantity, these simulations assume that the Ne/Nc ratio is 1:1 and is therefore known and accurate for all simulated populations. We use this assumption because previous work already showed that converting census size (Nc) with an incorrect Ne/Nc ratio can generate errors in Ne (Mastretta-Yanes et al., 2024) and therefore distort the Ne>500 indicator. By using this assumption, we can focus on the underexplored aspects of monitoring using the indicator: frequency, intensity, and data quality. We remind readers that when using the indicator in practice, the Ne/Nc ratio should only be applied to taxa for which it is known with a high degree of confidence, otherwise a range of Ne/Nc values should be tested.

2.1. Simulating population census data

2.1.1. Simulating the "true" population trend

For each scenario described below, we simulated 12 populations of a virtual species (based on vertebrate population dynamics; see Supplementary Material S1) from a starting census size $N_{t=1}$ through time following:

$$N_{true,\ t} = rN_{true,\ t\text{-}1} + \varepsilon_{\text{process}}$$

Where N_{true} is the population's census size, t is the time step, r is the annual growth rate, and ε is process error, which represents interannual population variability arising from demographic and environmental stochasticity. We simulated three directions of biodiversity change by changing the populations' average annual growth rate (r), which remained constant over the full 25 year period: (1) declining to under Ne 500, (2) stable at Ne 500, or (3) increasing to above Ne 500 (Table 1).

To evaluate the indicator's performance when applied to a mix of declining and growing trends, we simulated scenarios of conservation investment with three outcomes: 25%, 50%, or 75% of populations under Ne 500 recover to Ne>500 within 25 years (See Supplementary Material S2 for details and results). In these simulations, all populations were set to decline unless they benefit from conservation investment and each of these recovers at a different pace, with some recovering in 5 to 10 years and some recovering more slowly (>20 years) (Fig. S2.1). Because recovery was set to occur within 25 years, these simulations are more representative of species with relatively fast generation times, where enough generations occur within 25 years to increase the population's genetic diversity. However, observer error would have a similar effect on the indicator's reliability for species with slower generation times, though their recovery would need to be monitored over a longer timeline.

2.1.2. Simulating the "observed" population trend

We then simulated the "observed" population trend, which is the population census size data collected through a simulated monitoring scenario (Fig. S1). The observed population trend departs from the true population trend simulated above due to observer error and monitoring frequency (Fig. 1). In the following steps, population census size (Nc) is converted into effective size (Ne) with an Ne/Nc ratio of 1:1, under the assumption that this ratio is correct and known (but see Fig. S11: this conversion should only be done when this ratio is known with confidence).

2.1.2.1. Observer error

Detecting genuine variations in population size can be difficult, depending on the population's variability, size, time series length, and observer error (Clark & Bjørnstad, 2004; McCain et al., 2016; Arkilanian et al., 2020). Observer error introduces variability in population sizes due to variation in detection, sampling effort, observer skill, and measurement precision, among other factors (Hovestadt & Nowicki, 2008; Farmer et al., 2012). For example, indirect survey methods (e.g. tracks, pellets, etc.) may be more accurate for species that occupy large areas and are therefore more difficult to detect, compared to direct count methods (e.g. ground or aerial surveys) that can have higher observer error (Ahrestani et al., 2013).

It is then important to test how observer error affects the indicator's power detect a trend, so monitoring schemes can be tailored for indicator reporting and to allocate resources to manage the trade-off between census precision and sampling effort. To simulate observer error during the monitoring process, we randomly sampled each observed population's size at each time step from a Normal distribution:

$$\mathrm{Ne}_{obs} \sim N(\mathrm{Ne}_{true}, \sigma^2_{obs})$$

Where Ne_{obs} is the observed value of the population's size at time t, sampled from a Normal distribution with a mean of Ne_{true} (the population's true population size at time t without observer error) and a variance of σ^2_{obs} (observer error). We set three levels of observer error: low ($\sigma^2_{obs} = 0.05$), medium ($\sigma^2_{obs} = 0.1$), and high ($\sigma^2_{obs} = 0.2$). These simulations assume that observer error is unbiased for simplicity, though observer error is rarely random (Freckleton et al. 2006; Fitzpatrick et al. 2009; Dambly et al. 2020). The effects of biased observer error, where population size is systematically over- or under-estimated, on the Ne>500 indicator are explored in Supplementary Material (Fig. 8-9).

2.1.2.2. Monitoring frequency

We then defined five scenarios of monitoring frequency, ranging from monitoring all populations annually to every eight years (Table 2). Importantly, the value of each population's size is kept as the last observed value in the absence of new information during the time steps without monitoring (Fig.S1).

2.2. Performance testing

We tested indicator performance with 1000 randomisations of each set of 12 populations in the trend direction, level of observer error, and monitoring scenarios described above.

2.2.1. Step 1: Accurately detecting whether a population is above or below the Ne 500 threshold

The first step to calculate the Ne>500 indicator is to classify each population's observed effective population size as above or below the Ne 500 threshold (Fig. 2). This step introduces a potential source of error: if populations are misclassified with a systematic bias, the indicator will be biased. To evaluate how observer error and monitoring frequency influence the threshold success rate, we counted the number of time steps at which each population's observed position above/below the Ne 500 threshold matched its expected position given its simulated "true" population size (Fig. 1). We reported the success rate for this thresholding step as the proportion of correctly classified observations for all populations.

2.2.2. Step 2: Accurately summarising population trends at the species level

The second step for calculating the Ne>500 indicator is to count the proportion of monitored populations above the Ne 500 threshold, bound between 0 and 1, at the species level. To evaluate the indicator's accuracy, we measured the difference between the observed indicator value from the expected indicator value at each time step and in each scenario, reported here as the mean deviation with its standard deviation in each scenario. The observed indicator value was based on monitored populations that were subjected to observer error and different lags in monitoring, while the expected value is the "true" indicator based on annual population sizes without observer error.

2.3. Assessing detection power

2.3.1. How many populations need to be monitored to accurately represent a species?

To determine the proportion of populations that need to be monitored for an acceptably accurate species-level indicator, we generated 10 species with a range of 5 to 100 populations. We define the indicator to be "acceptably accurate" when it consistently falls within 10% of the expected "true" value throughout the entire 25-step time series. However, a 10% error rate can still have serious implications for conservation and should be redefined for specific applications, particularly if a species has few populations and/or requires a higher degree of confidence to ensure that conservation interventions are effective.

Each simulated population's starting size was pulled from a uniform distribution between 250 and 750 individuals, and its growth rate was drawn from a Normal distribution (mean = 1, sd

= 0.2). For each species, we simulated 1000 randomisations of the expected "true" trend and the "observed" trend with three levels of observer error (low = 0.05, medium = 0.1, and high = 0.2) and four monitoring frequencies (1 year, 2 years, 4 years, and 8 years). We calculated the indicator with 1000 randomisations for subsamples ranging from 1 to the total number of populations for each simulated species.

For each subsample, we measured the deviation between the expected "true" value and the "observed" value as above. We then evaluated the minimum subsample of populations needed to reach "acceptable accuracy", that is, for the species-level indicator to always lie within 10% of the expected "true" value for the given species.

2.3.2. How many species need to be monitored to accurately represent a country?

As with any summary metric, we expect the Ne>500 indicator to be more accurate (i.e., to better represent the country's trend distribution) when more species are included, especially if their trends are similar to the "true" average trend (i.e., less variation among species). To determine the number of species to monitor to ensure an "acceptably accurate" country-level indicator, where "acceptable accuracy" was defined as above, we used both empirical and simulated data. We first randomly subsampled Ne>500 indicator values calculated from empirical data for 583 species from Mastretta-Yanes and da Silva et al. (2024), representing a monitoring program that covers a portion of a country's species richness. For each subsample, we calculated the deviation between the estimated value (the indicator for a subsample of species) and the true value (the indicator for the full species list) for a range of species richness levels between 10 and 583 species, each for 10000 randomisations. We report the upper confidence limit (mean error + standard deviation) because this is a more conservative estimation of the risk of error. We then applied the same random subsampling approach to an extended empirical dataset (nearly 6000 species) with the same distribution as the data from Mastretta Yanes and da Silva *et al.* (2024) (Fig. S6). Finally, we subsampled from 5000 simulated species, with parameters similar to those in Table 1 (except starting population size and growth rate were pulled from distributions rather than set values; see Supplementary Material S3).

However, the selection of populations and species for long-term monitoring programs is rarely random (Yoccoz et al., 2001). In addition to common taxonomic and geographic biases (Dove et al., 2023), site-selection biases towards high-abundance sites artificially inflate the likelihood of monitoring declining trends (Fournier et al., 2019). We simulated biased sampling where species were weighted based on whether their populations were above or below Ne = 500. To simulate a bias towards low-Ne species, species with low values of the Ne < 500 indicator were twice as likely to be selected (and vice-versa for biased selection towards high-Ne species; see Supplementary Material Fig. S10).

3. Results

3.1. Step 1: Accurately detecting whether a population is above or below the Ne 500 threshold

We first evaluated the success rate when populations are classified as above or below the Ne 500 threshold (Fig. 2). We found that this threshold step is highly successful (82.3% to 97% success rate) in declining and growing populations in all monitoring scenarios with random observer error (Fig. 3). Populations that were stable at Ne 500 were classified the least successfully in all scenarios, ranging from 55.7% success under long-term monitoring with high observer error to 69.5% success under annual monitoring with low observer error.

When observer error is biased, these threshold errors are more likely in growth and decline scenarios (Fig. S8): if Ne is consistently overestimated, it is harder to detect when populations decline below the Ne 500 threshold (68.5% minimum success rate). If Ne is underestimated, the indicator can fail to detect growth past the Ne 500 threshold (74.1% minimum success rate).

Annual and short-term monitoring scenarios yielded almost identical success rates in all scenarios (Fig. 3, Table S1). Even under the long-term monitoring scenario and the highest level of observer error, declining and growing populations were classified successfully 82.8% and 82.3% of the time, respectively (Table S1).

3.2. Step 2: Accurately summarising population trends at the species level

Next, the population-level trends are aggregated to the species level. Overall, the observed indicator value tracks the true indicator value fairly closely when populations are monitored every one (1.2 ± 15%) to two years (0.8 ± 15%), particularly when observer error is low (annual: 0.8 ± 11%; short-term: 0.4 ± 11%) to moderate (annual: 0.9 ± 14%; short-term: 0.5 ± 14%) (Fig. 4a). The indicator is, on average, not biased towards over or underestimating the expected value: the median difference between the observed and the true indicator values was consistently zero across scenarios of decline and growth (Fig. 4). The indicator was typically less accurate when all populations were stable around Ne 500 (Fig. 4a), as it inherited the threshold errors from Step 1. When populations are stable at Ne 500, the species-level indicator ranged from best-case accuracy with annual monitoring and low observer error (1.9 ± 16% deviation) to worst-case accuracy from long-term monitoring with high observer error: (-12% ± 28% deviation).

The species-level indicator was also less accurate in slower monitoring scenarios (mixed, mid-term, and long-term). On average, the indicator deviated by -7.6 ± 18% (low observer error in growing trends) to 16 ± 22 % (high observer error in declining trends) from the true expected value under a long-term monitoring scenario, compared to a 0.2 ± 6 % to 2 ± 19 % deviation with an annual monitoring scenario (Fig. 4a). This drop in accuracy arises because longer lag times between monitoring events delay the correction of threshold errors in Step 1, which pushes the indicator further away from the expected value for longer periods of time.

3.3. How many populations need to be monitored to accurately represent a species?

Approximately half (and at least 25%) of a species' populations must be monitored to achieve a reasonably accurate estimate, which we defined here as the indicator being within 10% of the expected value (Fig. 5). If observer error is low to moderate, measuring 45% (CIs: 20-74%) to 49% (CIs: 22-81%) of a species' populations ensures 10% accuracy. If observer error is high (20% error around the observation), more populations must be monitored (58%, CIs: 27-100%) to reach the same level of accuracy. This is true regardless of the total number of populations in the range of our simulations, which range from 5 to 100 populations.

3.4. How many species need to be monitored to accurately represent a country?

It is possible to achieve a relatively low risk of error (5%) by monitoring a fraction of the targeted species' populations, depending on the targeted species' richness (Fig. 6). Assuming this goal, when the indicator is applied to a group of less than 100 species, almost 60% of species should be monitored, while roughly 30% of species should be monitored to represent groups of 200 species, and roughly a quarter (23%) of species when applied to groups of 300 to 500 species. When species richness is much higher (1000 to 5000 species), based on our results using an extended empirical dataset (details in Supplementary Material S3), approximately 8% of species should be monitored for a 5% risk of error (Fig. S6). Monitoring more species is of course recommended whenever possible but may not yield large gains in accuracy all else being equal (Fig. 6a). For example, achieving 97.5% accuracy will approximately double the monitoring effort in most cases, for example 14-15% of species (between 280 and 700 species) for the last example of a species pool of 2000 to 5000 species (Fig. S6). Simulations (up to 5000 species as described in Supplementary Material S3) similarly show that the first 25% of monitored species yield the greatest gains in accuracy (Fig. S7), depending on the total species richness the indicator is intended to represent. It is however important to note that these guidelines assume annual monitoring and are therefore a minimum: monitoring programs should aim for a larger representation of the targeted species if monitoring is slower or less precise.

These recommendations also assume that species are randomly represented in the sample data, though population census data typically represent a biased sample of species (Yoccoz et al., 2001; Fournier et al., 2019; Moussy et al., 2022; Dove et al., 2023). In this case, more species (74-87% of the pool) need to be monitored if species selection is systematically biased towards species with low or high Ne (Supplementary Fig. S10).

4. Discussion

An essential property of biodiversity indicators is that they summarize complex biodiversity change into simple, accurate and useful numbers (Jones et al., 2011; Stevenson et al., 2021). Robust testing with simulations and real data is essential to determine whether indicators accomplish this, and to establish their data quality and quantity requirements (Christie et al., 2020; Stevenson et al., 2021; Hébert et al., 2025). We used simulations and empirical data to test the Ne>500 indicator's performance in scenarios of biodiversity change depending on the quality and quantity of its input data. We find that the Ne>500 indicator is a reliable indicator of population changes (estimated by the proportion of populations with an effective size >500), except in the unlikely case that all populations are consistently stable around the Ne = 500 threshold.

To reliably detect change, a useful rule of thumb is to monitor or include data for at least 40% of a species' populations and at least 8% of large species pools (several thousand species), at least 23% of moderate species pools (300 to ~600 species) or 56% of small species pools (<100 species). This represents less than 100 to about 350 species, depending on the species pool. If low-Ne or high-Ne populations are overrepresented in the sampled pool, up to 74-87% of the species pool should be sampled. In addition to offering practical guidance for the Ne>500 indicator's data requirements, we suggest that our simulation-based approach can guide further performance testing to establish confidence in other biodiversity indicators, including those for the CBD GBF.

4.1. The performance of the Ne>500 indicator

We tested how data quantity and quantity impact the Ne>500 indicator's accuracy and detection power and found that, for a reliable species-level indicator, (1) monitoring more frequently is more important than monitoring more accurately, and (2) monitoring accuracy is most important when populations are likely to cross the Ne 500 threshold.

Policymakers will likely be reassured that the indicator detects both growing and declining population trends well, though with slightly weaker performance in detecting declines. The indicator actually performs best when applied to a mix of growing and declining populations, as we could expect in real datasets (Supplementary Material S2). The indicator's robustness is partly due to the Ne 500 threshold, because it is only likely to "miss the mark" when populations are crossing the threshold (Fig. S3 and S4). Threshold errors impact the indicator mostly when observer error is high and when monitoring is slower, particularly if populations are stable and fluctuating around Ne 500 (Fig. 3a). Though this "Stable at Ne 500" scenario is unlikely in real populations, it highlights the indicator's main source of error: random fluctuations in Ne around 500, which can arise due to stochastic population dynamics and observer error, can falsely push a population above or below the Ne 500 threshold. The impacts of these errors might be amplified if they are correlated among populations through time, as they can quickly accumulate when trends are aggregated (Hébert & Gravel, 2023). When population sizes are far from the threshold during most of the time series or definitively cross the threshold, as in the decline and growth scenarios, error rates are low. Frequent monitoring reduces the impact of threshold errors: as the indicator's value does not depend on previous years, the indicator is corrected as soon as the next monitoring event occurs if the new observation is more accurate. Threshold errors are therefore unlikely to distort the indicator unless Ne hovers around the threshold value (i.e., Ne 500) for long periods of time, or unless observer error is high enough to risk falsely pushing the observed Ne across the threshold.

However, threshold errors can still have serious implications for the conservation of genetic diversity, depending on whether populations incorrectly falls below or above Ne 500 and how these errors accumulate when populations are aggregated into the Ne>500 indicator (Supplementary Material Fig. S2). A false positive, where a population below Ne 500 is falsely classified as being above the threshold, may give the false impression that a population is healthy or has recovered (i.e., a “false green flag”) and can artificially ameliorate the indicator. A false negative, where a population above Ne 500 is classified below the threshold, may instead be a “false alarm”, which artificially pulls the indicator down. Though both cases are errors, the first case has more serious implications for conservation: false green flags can drive the indicator to understate declines, prematurely slowing or stopping conservation resources that are still needed to maintain genetic diversity. In the second case, false alarms can artificially pull the indicator downwards and drain conservation resources needlessly, but genetic diversity will not actually be damaged because of the error. Knowing the risk of false positives and false negatives and how they accumulate at the indicator level is therefore crucial to understanding the ecological implications of making misinformed decisions based on the indicator.

4.2. Data quality and quantity requirements for a reliable indicator

Here, our simulations assume that Ne is derived from census size data (Nc) via a known Ne/Nc ratio, because there are currently very few species for which there is regularly repeated, multi-population sampling of DNA data to estimate Ne for biodiversity reporting, even in high income countries (Posledovich et al., 2021; Mastretta-Yanes et al., 2024). Using an incorrect Ne/Nc ratio could be a significant error when using census data (Mastretta-Yanes et al., 2024), which we did not explore here, because we aimed to isolate the effects of observer error, population growth trajectory, and monitoring scenario. Until at least 2030, The Ne>500 indicator is likely to rely on population census data. However, relying on census data alone will not be sufficient to understand the drivers and mechanisms of genetic diversity loss, and therefore to act effectively to halt these losses. Given the importance of genetic diversity in the Global Biodiversity Framework, as demonstrated by the Ne>500 indicator being set as a headline indicator (CBD 2022), investing in DNA based monitoring of a wide range of species is critical to track genetic diversity trends (Pearman et al., 2011; Shaw et al., 2025). We note that our results about the frequency of monitoring could be extrapolated to DNA-based monitoring, notably by substituting the Ne-estimation-error rate (as established in some DNA studies; e.g. (Gilbert & Whitlock, 2015)) for the observer rate used here.

Monitoring lags and observer errors interact and may have different impacts depending on population sizes and growth rates, so all these factors are important considerations to ensure reliable indicator reporting. Annual or bi-annual sampling designs may be most appropriate for populations that are likely to grow or decline past Ne 500 within the targeted timeframe, while larger and more stable populations may only require monitoring every 4 to 8 years. However, lags from slower monitoring introduce an asymmetric bias: the indicator responds too slowly to change and is thus overly optimistic when populations decline, and overly pessimistic when populations are growing. Though the magnitude of this bias is similar between the decline and growth scenarios, it is more dangerous for conservation and management when populations are declining. More generally, it may be worthwhile to invest in determining how accurate observations really are early in a monitoring design (due to detection issues, sampling effort, logistical constraints, etc.) to anticipate and mitigate the risk of making errors in the indicator (Farmer et al., 2012; Isaac et al., 2014; Bennett et al., 2024). For example, investing in more frequent monitoring when there is a high likelihood of error can help to update the indicator and therefore minimize the potential impacts of large observer errors through time.

Ultimately, monitoring decisions should be tailored to the focal species and their populations. First, assessing a population's variability is key to predicting how often and how severely it could cross the Ne 500 threshold, and therefore how to monitor it. Specifically, populations that vary a lot from year to year may require more monitoring resources to detect the population's true size (Gibbs et al., 1998; McCain et al., 2016; Wauchope et al., 2019). Monitoring frequency should also match the populations' dynamics in a meaningful way to detect changes in genetic diversity (Schwartz et al., 2007): for example, species with fast life histories, in which genetic drift occurs rapidly, may require more frequent monitoring. Second, the minimum sample size of species and populations may differ across taxonomic groups and geographic regions. For example, some harder-to-detect taxonomic groups may require more sampling effort (Bennett et al., 2024). Regions with higher species diversity may also require larger sample sizes. For example, the Living Planet Index, an indicator of relative change in population sizes, requires larger sample sizes in the more biodiverse global south than in the global north, as well as for groups like fishes and reptiles and amphibians (Dove et al., 2023).

In terms of number of species to monitor, we found that a general rule of thumb is to monitor less than 100 to about 350 species across the species pools we tested (100 to 5000 species). This is on the same scale, though somewhat higher than the sample sizes used in a recent pilot study to report the indicator for nine countries, which aimed to include 50 to 100 species per country (Mastretta-Yanes et al., 2024). The finding that monitoring approximately

45% to 60% of populations is sufficient is important to guide future indicator reporting as well, as many species included in the pilot study did not have available data for all their populations (Mastretta-Yanes et al., 2024).

Though these recommendations can help to guide the scope of a monitoring program or the collection of existing data, the choice of which species and populations to include must still be carefully considered. This choice includes balancing conservation, political, and logistical considerations (McDonald-Madden et al., 2010; Bal et al., 2018), which can determine whether the aim is to monitor a random sample of the targeted biodiversity, the most ecologically vulnerable species and populations, or the populations that are under recovery plans (Hvilsom et al., 2022). Our recommendations can help to design monitoring programs for tracking population-level biodiversity changes, but they should also be tailored to more specific needs whenever possible and re-evaluated with empirical data as it accumulates for the genetic diversity indicator. Though it may be necessary to leverage existing data at first, we need to move towards strategically designing monitoring programs that provide the required data quality and quantity for robust indicators (Gonzalez et al., 2023). This includes filling gaps and reduce biases in existing data for rapid change detection with indicators like Ne>500, such as targeting sites and species that are expected to change and using adaptive sampling strategies to adjust monitoring towards updated conservation priorities. In addition to being strategic, monitoring must be globally coordinated, as envisioned by the proposed global biodiversity observing system (GBiOS), to inform and motivate global action (Gonzalez et al., 2023).

4.3. What this means for other biodiversity indicators

To understand the performance of other biodiversity indicators of the GBF, the steps outlined here could be applied to other indicators (including "Proportion of Populations Maintained") under realistic scenarios of biodiversity change and monitoring. Importantly, our findings may generally apply to other biodiversity indicators that track changes in a categorical or binary value. For example, the Red List Index tracks changes in species' extinction risk status through time, which ultimately stems from threshold values for many of its criteria (rate of decline, population sizes, numbers of populations, etc.). However, species' IUCN Red List statuses regularly take 10 years or longer to be re-evaluated (Rondinini et al., 2014; Cazalis et al., 2024). There are likely numerous species which are currently more or less threatened than their Red List status indicates, which can introduce error into the Red List Index until these data are updated and until the criteria thresholds are definitively crossed. Going forward, performance testing is a key step to design monitoring programs that connect to indicator needs, to ensure that these can be reliably reported to track progress towards global biodiversity targets such as the Kunming-Montreal GBF. The CBD is establishing Technical and Scientific Cooperation Support Centres around the world, and meanwhile, GEO BON and other entities are preparing to leverage thousands of volunteer scientific experts to support Parties in their reporting. Studies like this one are a key part of that support, and of the envisioned Global Biodiversity Observation System (GBiOS).

5. Conclusion

Biodiversity indicators translate raw data into decisions that shape the future of biodiversity, but their influence is limited by how much confidence we place in them. This confidence can only come from clearly defining the conditions in which these indicators work as intended. As with many indicators, the Ne>500 indicator currently depends on data that is already available, which consists largely of population census data. We recommend that these data fit certain criteria to ensure the Ne>500 indicator can reliably report change: at least half of a species' populations should be monitored every 1 to 2 years, at least 40% of a species' populations should be included, and at least 8% and up to 86% of the species pool, depending on the size of the pool. Going forward, genetic monitoring programs are urgently needed to link this indicator more firmly to genetic diversity change, particularly given the Global Biodiversity Framework's emphasis on genetic diversity. Ultimately, performance testing must become standard practice to design effective monitoring programs and to establish confidence in the indicators we use to safeguard biodiversity.

Data and code availability statement: Code to generate, analyze, and visualize the simulated data in this study is accessible on GitHub (https://github.com/katherinehebert/Ne_scenarios). No data was collected for this study.

Acknowledgements

K.H. was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance Grant (Université de Sherbrooke, Biodiversité Québec). S.H. would like to thank the Walder Foundation Biota Award which partly financially supported this work. He would also like to thank the colleagues at the CBD Secretariat, GEO BON Secretariat, Coalition for Conservation Genetics, and IUCN Global Biodiversity Framework Task Force for discussions on this topic. We thank the Conservation Biology Lab at The Morton Arboretum and Paul Savary for insightful discussions and feedback on the manuscript.

Tables

Table 1. Parameters for simulations of three directions of population trends.

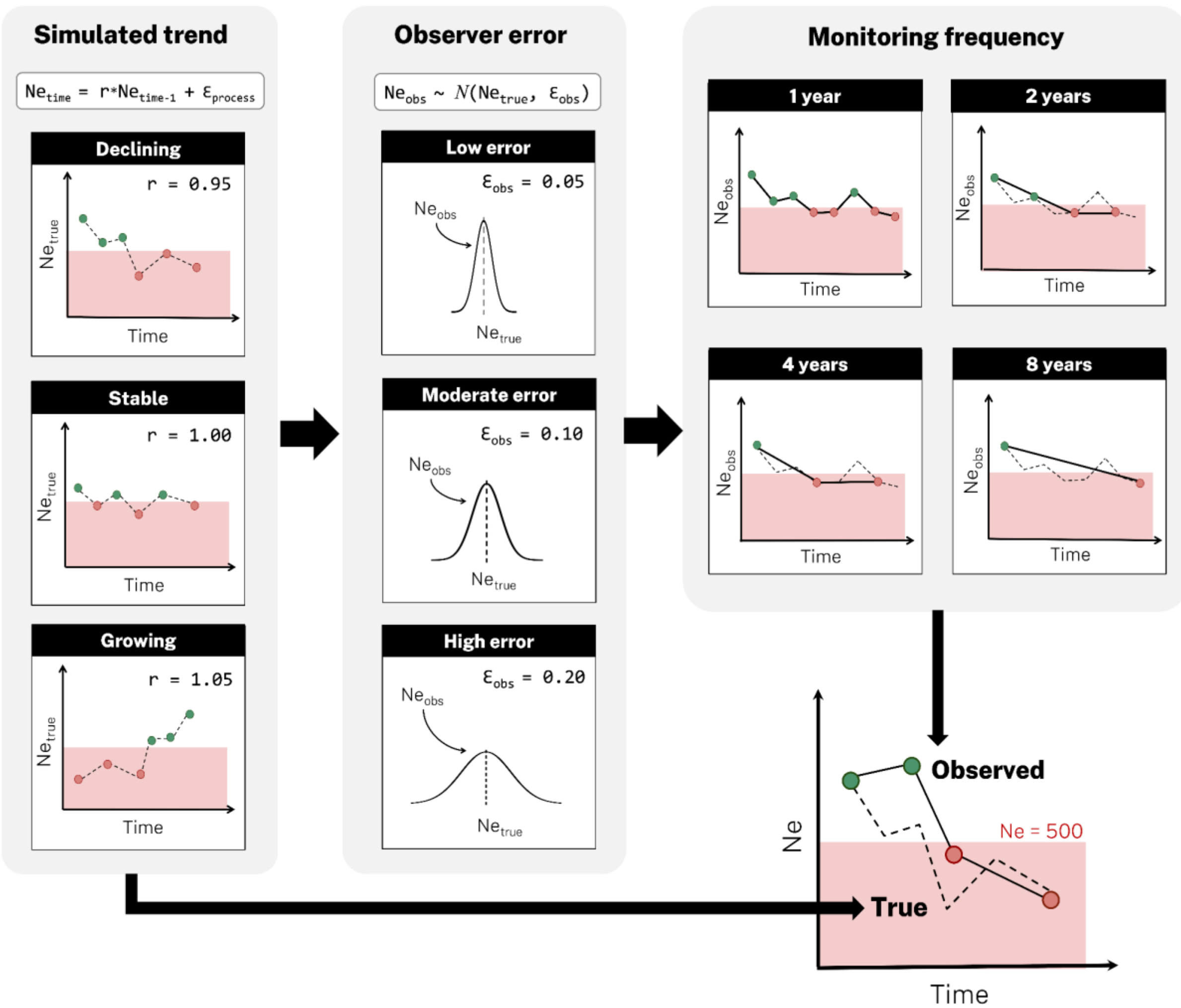
We grounded parameter selection in empirical data and in relevant values for situations of concern for conservation (more details in Supplementary Material S1).

		Population trends		
	symbol	**Decline**	**Stable at Ne 500**	**Growth**
Time steps	t	25	25	25
Number of populations	n	12	12	12
Initial population size	$N_{t=1}$	750	500	250
Annual growth rate	r	0.95	1.00	1.05
Process error	$\varepsilon_{\text{process}}$	0.01	0.01	0.01

Table 2. Monitoring scenarios which illustrate the proportion of populations monitored at different frequencies through time.

Monitoring scenario	**Monitoring frequency**			
	each year	**every 2 years**	**every 4 years**	**every 8 years**
Annual	100%			
Short-term	50%	50%		
Mixed	25%	25%	25%	25%
Mid-term			50%	50%
Long-term				100%

Figures



Simulated trend
Ne_time = r*Ne_time-1 + ε_process
Declining
r = 0.95
Ne_true
Time
Stable
r = 1.00
Ne_true
Time
Growing
r = 1.05
Ne_true
Time
Observer error
Ne_obs ~ N(Ne_true, ε_obs)
Low error
ε_obs = 0.05
Ne_obs
Ne_true
Moderate error
ε_obs = 0.10
Ne_obs
Ne_true
High error
ε_obs = 0.20
Ne_obs
Ne_true
Monitoring frequency
1 year
Ne_obs
Time
2 years
Ne_obs
Time
4 years
Ne_obs
Time
8 years
Ne_obs
Time
Observed
Ne = 500
Ne
True
Time

Figure 1. Simulating true and observed trends in effective population size (Ne). First, each population's true Ne trend is simulated from a constant annual growth rate (declining, stable, or growing) with some process error to represent stochastic population dynamics. This true Ne trend is then observed, where each Ne value is sampled from a Normal distribution centered on the true Ne value with three levels of observation error (low, moderate, high). Then, the observed Ne trends are monitored at different frequencies (every 1, 2, 4, or 8 years). For each population, we can then determine how observer error and monitoring frequency combine to result in an observed trend that differs from the true trend.

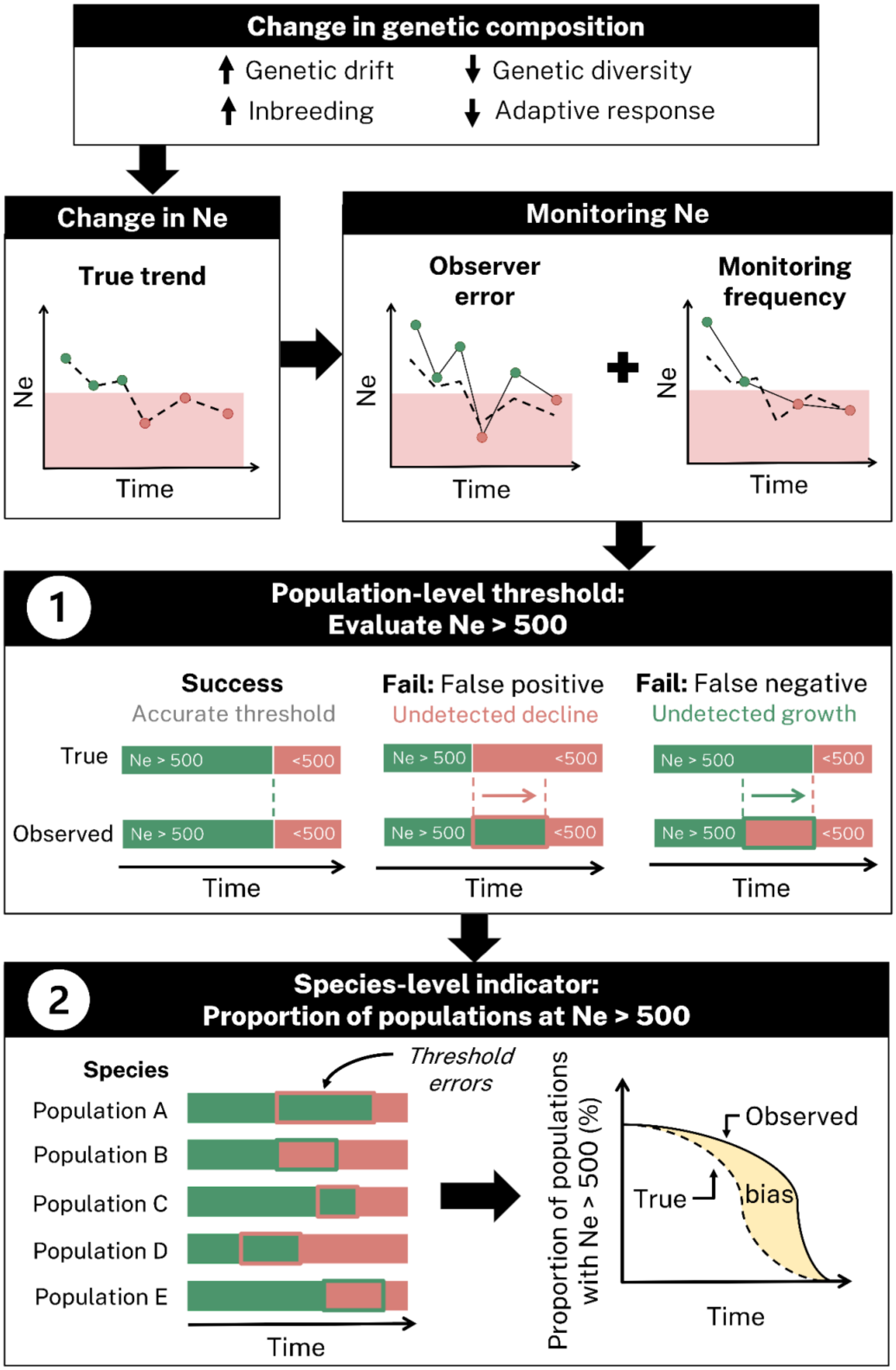

Change in genetic composition
Genetic drift
Genetic diversity
Inbreeding
Adaptive response
Change in Ne
True trend
Ne
Time
Monitoring Ne
Observer error
Monitoring frequency
1
Population-level threshold: Evaluate Ne > 500
Success
Accurate threshold
Fail: False positive
Undetected decline
Fail: False negative
Undetected growth
True
Observed
Ne > 500
<500
2
Species-level indicator: Proportion of populations at Ne > 500
Species
Population A
Population B
Population C
Population D
Population E
Threshold errors
Proportion of populations with Ne > 500 (%)
Observed
True
bias
Time

Figure 2. How monitoring can potentially introduce errors into the Ne > 500 indicator. First, changes in genetic composition are reflected in the change in Ne through time. When each population is monitored, observer error and monitoring frequency can distort the true Ne trend, resulting in an observed Ne trend. To calculate the indicator, Step 1 is a population-level threshold: each population is evaluated to identify years at which the population is below the Ne 500 threshold. At this step, observer error and monitoring frequency can combine to generate lags in the timing of the threshold being crossed, resulting in a threshold error. When populations are combined into a species-level indicator at Step 2, this threshold error can accumulate across populations at each time step, propagating these lags into the species-level indicator. These threshold errors then have the potential to generate a bias between the observed and true indicator trend.

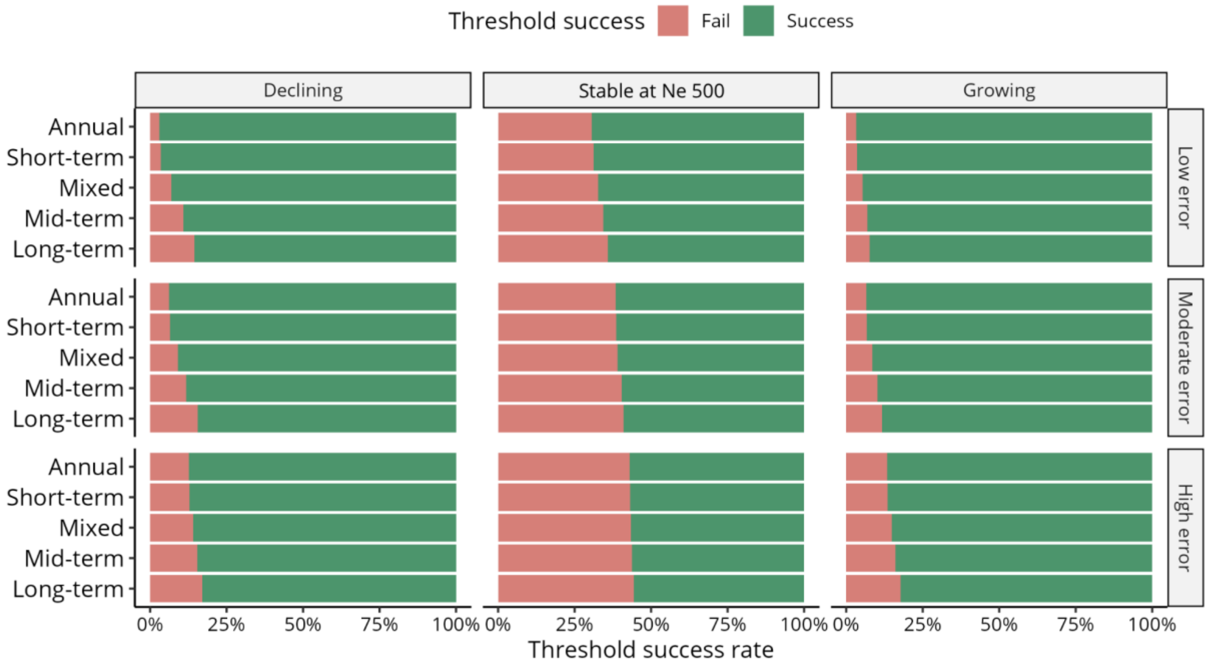


Figure 3. Success rate of the Ne > 500 threshold evaluation step in each scenario with 1000 randomisations. Scenarios in which the 12 populations of a species all follow the same average growth rate (declining, stable at Ne 500, growing). The green bar shows the success rate, i.e. the proportion of data points that were successfully classified above or below the threshold, while red shows the rate of failure in each scenario. Successes and failures are broken down into true and false positives/negatives in Fig. S2.

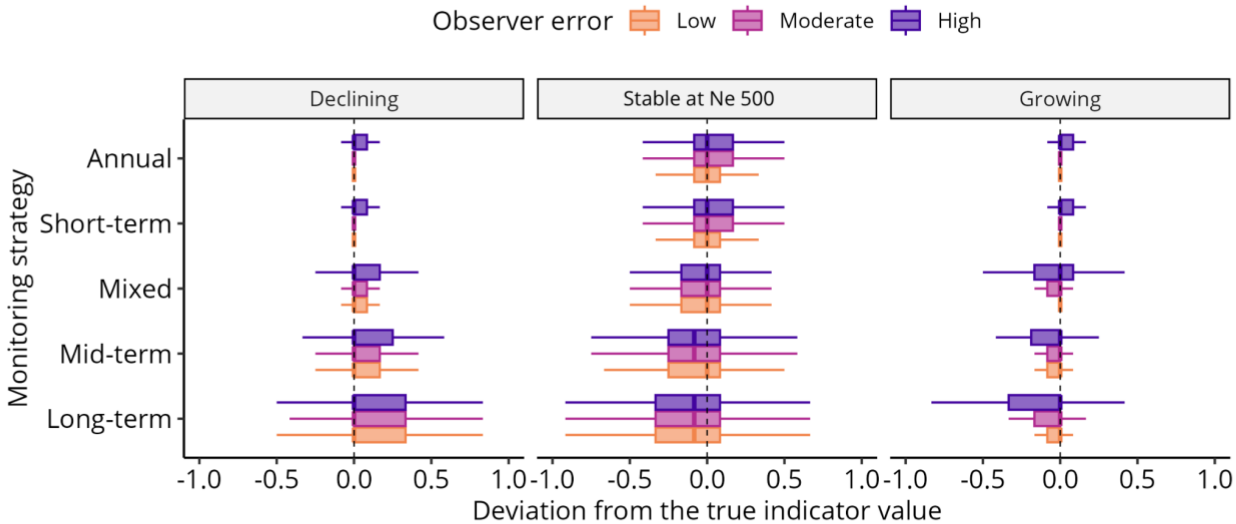


Figure 4. The species-level "Proportion of populations with an Ne > 500" indicator's sensitivity to observer error and monitoring scenario. Boxplots show the observed indicator's deviation from the true indicator trend in each scenario, over 1000 randomisations. Each panel shows a biodiversity change scenario in which the 12 populations of a species vary at the same annual growth rate (declining, stable at Ne 500, growing). Colours show three levels of observer error introduced into the observed population sizes that were used to calculate the observed indicator value.

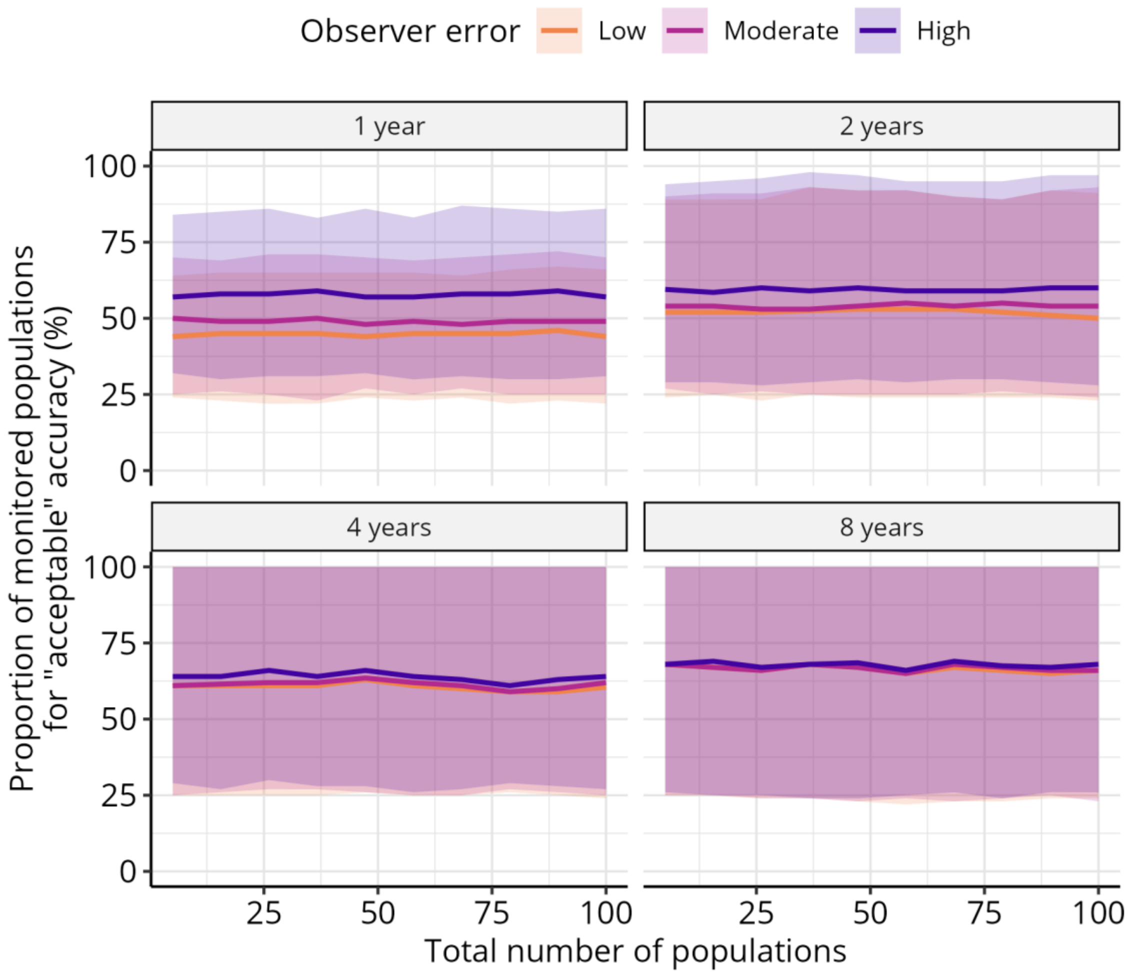


Figure 5. Proportion of populations that must be monitored for an "acceptably accurate" species-level indicator, depending on the total number of populations included in the species. "Acceptable accuracy" is defined here as an indicator that is consistently within 10% of the expected "true" value throughout the entire time series, evaluated over 1000 randomisations. Panels show different monitoring frequencies and colours show three levels of observer error (low = 0.05, medium = 0.1, and high = 0.2).

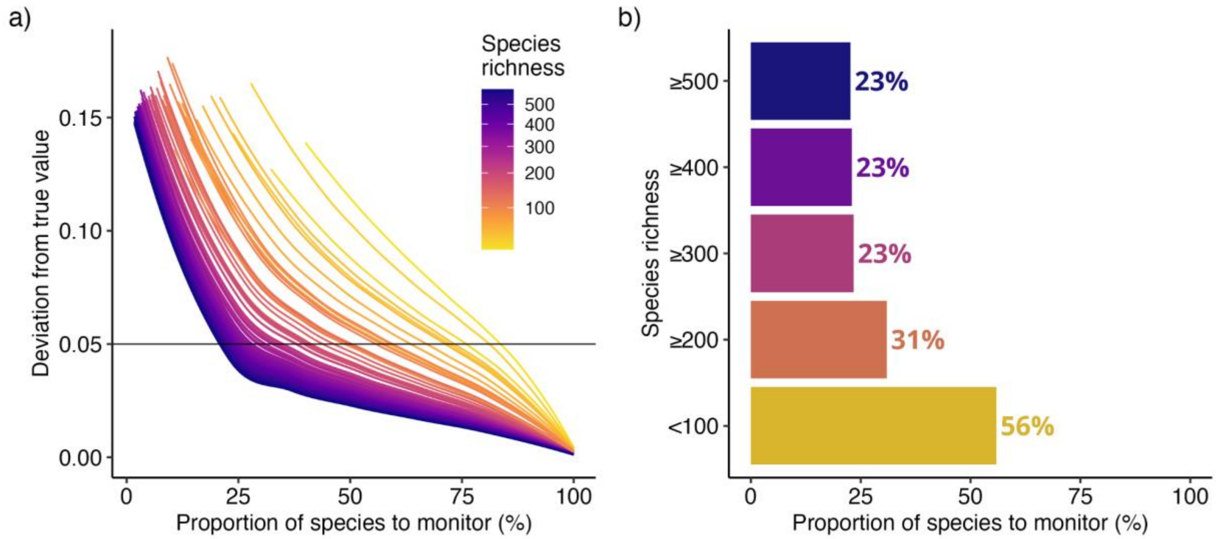


Figure 6. How many species to monitor for an accurate (5% error) country-level Ne>500 indicator relative to the country's full species pool, based on a subsampling approach with data from Mastretta-Yanes et al. (2024). (a) Deviation between the estimated value (the indicator for a subsample of species) and the true value (the indicator for the full species list) for a range of species richness levels between 10 and 583 species, each for 10000 randomisations. We show the upper confidence limit (mean error + standard deviation) because this is a more conservative estimation of the risk of error. The horizontal line shows the proportion of species that must be monitored to risk 5% or less error from the true country-level indicator value, summarised in panel (b). (b) Proportion of species to monitor to risk 5% or less error of the country-level Ne>500 indicator as shown by the horizontal line in panel (a). Note that these recommendations assume that species are randomly represented in the sample and that more species (up to 86% of the pool) need to be monitored if species selection is systematically biased towards species with high Ne (>500) or low Ne (<500) (Supplementary Fig. S10).

References

Affinito, F., Williams, J. M., Campbell, J. E., Londono, M. C., & Gonzalez, A. (2024). Progress in developing and operationalizing the Monitoring Framework of the Global Biodiversity Framework. *Nature Ecology & Evolution*, 1–9. https://doi.org/10.1038/s41559-024-02566-7

Ahrestani, F. S., Hebblewhite, M., & Post, E. (2013). The importance of observation versus process error in analyses of global ungulate populations. *Scientific Reports*, *3*(1), Article 1. https://doi.org/10.1038/srep03125

Allendorf, F. W., Hössjer, O., & Ryman, N. (2024). What does effective population size tell us about loss of allelic variation? *Evolutionary Applications*, *17*(6), e13733. https://doi.org/10.1111/eva.13733

Arkilanian, A. A., Clements, C. F., Ozgul, A., & Baruah, G. (2020). Effect of time series length and resolution on abundance- and trait-based early warning signals of population declines. *Ecology*, *101*(7). https://doi.org/10.1002/ecy.3040

Bal, P., Tulloch, A. I., Addison, P. F., McDonald-Madden, E., & Rhodes, J. R. (2018). Selecting indicator species for biodiversity management. *Frontiers in Ecology and the Environment*, *16*(10), 589–598. https://doi.org/10.1002/fee.1972

Bennett, J., Edwards, B., Bergman, J., Binley, A., Buxton, R., Hanna, D., Hanson, J., Hudgins, E., Karimi, S., Raymond, C., Robichaud, C., & Rytwinski, T. (2024). How ignoring detection probability hurts biodiversity conservation. *Frontiers in Ecology and the Environment*. https://doi.org/10.1002/fee.2782

Cazalis, V., Di Marco, M., Zizka, A., Butchart, S. H. M., González-Suárez, M., Böhm, M., Bachman, S. P., Hoffmann, M., Rosati, I., De Leo, F., Jung, M., Benítez-López, A., Clausnitzer, V., Cardoso, P., Brooks, T. M., Mancini, G., Lucas, P. M., Young, B. E., Akçakaya, H. R., … Santini, L. (2024). Accelerating and standardising IUCN Red List assessments with sRedList. *Biological Conservation*, *298*, 110761. https://doi.org/10.1016/j.biocon.2024.110761

CBD/SBSTTA/26/2. (2024). Monitoring framework for the Kunming-Montreal Global Biodiversity Framework. *Convention on Biological Diversity*, *26/2*, 1–29.

Cerini, F., Childs, D. Z., & Clements, C. F. (2023). A predictive timeline of wildlife population collapse. *Nature Ecology & Evolution*, *7*(3), Article 3. https://doi.org/10.1038/s41559-023-01985-2

Christie, A. P., Amano, T., Martin, P. A., Petrovan, S. O., Shackelford, G. E., Simmons, B. I., Smith, R. K., Williams, D. R., Wordley, C. F. R., & Sutherland, W. J. (2020). Poor availability of context-specific evidence hampers decision-making in conservation. *Biological Conservation*, *248*, 108666. https://doi.org/10.1016/j.biocon.2020.108666

Clark, J. S., & Bjørnstad, O. N. (2004). Population time series: Process variability, observation errors, missing values, lags, and hidden states. *Ecology*, *85*(11), 3140–3150. https://doi.org/10.1890/03-0520

Dove, S., Böhm, M., Freeman, R., McRae, L., & Murrell, D. J. (2023). Quantifying reliability and data deficiency in global vertebrate population trends using the Living Planet Index. *Global Change Biology*, *00*(n/a), 1–17. https://doi.org/10.1111/gcb.16841

Farmer, R. G., Leonard, M. L., & Horn, A. G. (2012). Observer Effects and Avian-Call-Count Survey Quality: Rare-Species Biases and Overconfidence. *The Auk*, *129*(1), 76–86. https://doi.org/10.1525/auk.2012.11129

Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences*, *111*(Supplement 4), 13664–13671. https://doi.org/10.1073/pnas.1317504111

Flather, C. H., Wilson, K. R., Dean, D. J., & McComb, W. C. (1997). Identifying Gaps in Conservation Networks: Of Indicators and Uncertainty in Geographic-Based Analyses. *Ecological Applications*, *7*(2), 531–542. https://doi.org/10.1890/1051-0761(1997)007[0531:IGICNO]2.0.CO;2

Fournier, A. M. V., White, E. R., & Heard, S. B. (2019). Site-selection bias and apparent population declines in long-term studies. *Conservation Biology*. https://doi.org/10.1111/cobi.13371

Frankham, R. (2022). Evaluation of proposed genetic goals and targets for the Convention on Biological Diversity. *Conservation Genetics*, *23*(5), 865–870. https://doi.org/10.1007/s10592-022-01459-1

Franklin, I. R., & Frankham, R. (1998). How large must populations be to retain evolutionary potential? *Animal Conservation*, *1*(1), 69–70. https://doi.org/10.1111/j.1469-1795.1998.tb00228.x

Gibbs, J. P., Droege, S., & Eagle, P. (1998). Monitoring populations of plants and animals. *BioScience*, *48*(11), 935–940. https://doi.org/10.2307/1313297

Gilbert, K. J., & Whitlock, M. C. (2015). Evaluating methods for estimating local effective population size with and without migration. *Evolution*, *69*(8), 2154–2166. https://doi.org/10.1111/evo.12713

Gonzalez, A., Vihervaara, P., Balvanera, P., Bates, A. E., Bayraktarov, E., Bellingham, P. J., Bruder, A., Campbell, J., Catchen, M. D., Cavender-Bares, J., Chase, J., Coops, N., Costello, M. J., Dornelas, M., Dubois, G., Duffy, E. J., Eggermont, H., Fernandez, N., Ferrier, S., … Wright, E. (2023). A global biodiversity observing system to unite monitoring and guide action. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-023-02171-0

Hébert, K., & Gravel, D. (2023). The Living Planet Index's ability to capture biodiversity change from uncertain data. *Ecology*, *104*(6), e4044. https://doi.org/10.1002/ecy.4044

Hébert, K., Jousse, M., Serrano, J., Karger, D. N., Blanchet, F. G., & Pollock, L. J. (2025). Five recommendations to fill the blank space in indicators at local and short-term scales. *Biological Conservation*, *302*, 111007. https://doi.org/10.1016/j.biocon.2025.111007

Hellmann, J. J., & Fowler, G. W. (1999). Bias, Precision, and Accuracy of Four Measures of Species Richness. *Ecological Applications*, *9*(3), 824–834. https://doi.org/10.1890/1051-0761(1999)009[0824:BPAAOF]2.0.CO;2

Hoban, S., Arntzen, J. A., Bruford, M. W., Godoy, J. A., Rus Hoelzel, A., Segelbacher, G., Vilà, C., & Bertorelle, G. (2014). Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. *Evolutionary Applications*, *7*(9), 984–998. https://doi.org/10.1111/eva.12197

Hoban, S., Campbell, C. D., da Silva, J. M., Ekblom, R., Funk, W. C., Garner, B. A., Godoy, J. A., Kershaw, F., MacDonald, A. J., Mergeay, J., Minter, M., O'Brien, D., Vinas, I. P., Pearson, S. K., Pérez-Espona, S., Potter, K. M., Russo, I.-R. M., Segelbacher, G., Vernesi, C., & Hunter, M. E. (2021). Genetic diversity is considered important but interpreted narrowly in country reports to the Convention on Biological Diversity: Current actions and indicators are insufficient. *Biological Conservation*, *261*, 109233. https://doi.org/10.1016/j.biocon.2021.109233

Hoban, S., da Silva, J. M., Hughes, A., Hunter, M. E., Kalamujić Stroil, B., Laikre, L., Mastretta-Yanes, A., Millette, K., Paz-Vinas, I., Bustos, L. R., Shaw, R. E., Vernesi, C., & the Coalition for Conservation Genetics. (2024). Too simple, too complex, or just right? Advantages, challenges, and guidance for indicators of genetic diversity. *BioScience*, *74*(4), 269–280. https://doi.org/10.1093/biosci/biae006

Hoban, S., da Silva, J. M., Mastretta-Yanes, A., Grueber, C. E., Heuertz, M., Hunter, M. E., Mergeay, J., Paz-Vinas, I., Fukaya, K., Ishihama, F., Jordan, R., Köppä, V., Latorre-Cárdenas, M. C., MacDonald, A. J., Rincon-Parra, V., Sjögren-Gulve, P., Tani, N., Thurfjell, H., & Laikre, L. (2023). Monitoring status and trends in genetic diversity for the Convention on Biological Diversity: An ongoing assessment of genetic indicators in nine countries. *Conservation Letters*, *16*(3), e12953. https://doi.org/10.1111/conl.12953

Hovestadt, T., & Nowicki, P. (2008). Process and measurement errors of population size: Their mutual effects on precision and bias of estimates for demographic parameters. *Biodiversity and Conservation*, *17*(14), 3417–3429. https://doi.org/10.1007/s10531-008-9426-0

Hvilsom, C., Segelbacher, G., Ekblom, R., Fischer, M. C., Laikre, L., Leus, K., O'Brien, D., Shaw, R., & Sork, V. (2022). *Selecting species and populations for monitoring of genetic* *diversity*. IUCN, International Union for Conservation of Nature. https://doi.org/10.2305/IUCN.CH.2022.07.en

Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, *5*(10), 1052–1060. https://doi.org/10.1111/2041-210X.12254

Jones, J. P. G., Collen, B. E. N., Atkinson, G., Baxter, P. W. J., Bubb, P., Illian, J. B., Katzner, T. E., Keane, A., Loh, J., & MCDONALD-MADDEN, E. V. E. (2011). The why, what, and how of global biodiversity indicators beyond the 2010 target. *Conservation Biology*, *25*(3), 450–457.

Leung, B., & Gonzalez, A. (2024). Global monitoring for biodiversity: Uncertainty, risk, and power analyses to support trend change detection. *Science Advances*, *10*(7), eadj1448. https://doi.org/10.1126/sciadv.adj1448

Lynch, M., & Lande, R. (1998). The critical effective size for a genetically secure population. *Animal Conservation*, *1*(1), 70–72. https://doi.org/10.1111/j.1469-1795.1998.tb00229.x

Mastretta-Yanes, A., da Silva, J. M., Grueber, C. E., Castillo-Reina, L., Köppä, V., Forester, B. R., Funk, W. C., Heuertz, M., Ishihama, F., Jordan, R., Mergeay, J., Paz-Vinas, I., Rincon-Parra, V. J., Rodriguez-Morales, M. A., Arredondo-Amezcua, L., Brahy, G., DeSaix, M., Durkee, L., Hamilton, A., … Hoban, S. (2024). Multinational evaluation of genetic diversity indicators for the Kunming-Montreal Global Biodiversity Framework. *Ecology Letters*, *27*(7), e14461. https://doi.org/10.1111/ele.14461

McCain, C., Szewczyk, T., & Bracy Knight, K. (2016). Population variability complicates the accurate detection of climate change responses. *Global Change Biology*, *22*(6), 2081–2093. https://doi.org/10.1111/gcb.13211

McDonald-Madden, E., Baxter, P. W. J., Fuller, R. A., Martin, T. G., Game, E. T., Montambault, J., & Possingham, H. P. (2010). Monitoring does not always count. *Trends in Ecology & Evolution*, *25*(10), 547–550. https://doi.org/10.1016/j.tree.2010.07.002

Moussy, C., Burfield, I. J., Stephenson, P. J., Newton, A. F. E., Butchart, S. H. M., Sutherland, W. J., Gregory, R. D., McRae, L., Bubb, P., Roesler, I., Ursino, C., Wu, Y., Retief, E. F., Udin, J. S., Urazaliyev, R., Sánchez-Clavijo, L. M., Lartey, E., & Donald, P. F. (2022). A quantitative global review of species population monitoring. *Conservation Biology*, *36*(1), e13721. https://doi.org/10.1111/cobi.13721

Nicholson, E., Fulton, E. A., Brooks, T. M., Blanchard, R., Leadley, P., Metzger, J. P., Mokany, K., Stevenson, S., Wintle, B. A., Woolley, S. N. C., Barnes, M., Watson, J. E. M., & Ferrier, S. (2019). Scenarios and Models to Support Global Conservation Targets. *Trends in Ecology & Evolution*, *34*(1), 57–68. https://doi.org/10.1016/j.tree.2018.10.006

Nicholson, E., Watermeyer, K. E., Rowland, J. A., Sato, C. F., Stevenson, S. L., Andrade, A., Brooks, T. M., Burgess, N. D., Cheng, S.-T., Grantham, H. S., Hill, S. L., Keith, D. A., Maron, M., Metzke, D., Murray, N. J., Nelson, C. R., Obura, D., Plumptre, A., Skowno, A. L., & Watson, J. E. M. (2021). Scientific foundations for an ecosystem goal, milestones and indicators for the post-2020 global biodiversity framework. *Nature Ecology & Evolution*, *5*(10), 1338–1349. https://doi.org/10.1038/s41559-021-01538-5

Noss, R. F. (1990). Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conservation Biology*, *4*(4), 355–364. https://doi.org/10.1111/j.1523-1739.1990.tb00309.x

Pearman, P. B., Guisan, A., & Zimmermann, N. E. (2011). Impacts of climate change on Swiss biodiversity: An indicator taxa approach. *Biological Conservation*, *144*(2), 866–875. https://doi.org/10.1016/j.biocon.2010.11.020

Posledovich, D., Ekblom, R., & Laikre, L. (2021). Mapping and monitoring genetic diversity in Sweden. *The Swedish Environmental Protection Agency*, *Report 6959*.

Reed, D. H., & Hobbs, G. R. (2004). The relationship between population size and temporal variability in population size. *Animal Conservation*, *7*(1), 1–8. https://doi.org/10.1017/S1367943004003476

Robuchon, M., da Silva, J., Dubois, G., Gumbs, R., Hoban, S., Laikre, L., Owen, N. R., & Perino, A. (2023). Conserving species' evolutionary potential and history: Opportunities under the Kunming–Montreal Global Biodiversity Framework. *Conservation Science and Practice*, *5*(6), e12929. https://doi.org/10.1111/csp2.12929

Rondinini, C., Di Marco, M., Visconti, P., Butchart, S. H. M., & Boitani, L. (2014). Update or Outdate: Long-Term Viability of the IUCN Red List. *Conservation Letters*, *7*(2), 126–130. https://doi.org/10.1111/conl.12040

Rowland, J. A., Bland, L. M., James, S., & Nicholson, E. (2021). A guide to representing variability and uncertainty in biodiversity indicators. *Conservation Biology*, *35*(5), 1669–1682. https://doi.org/10.1111/cobi.13699

Santos-del-Blanco, L., Olsson, S., Budde, K. B., Grivet, D., González-Martínez, S. C., Alía, R., & Robledo-Arnuncio, J. J. (2022). On the feasibility of estimating contemporary effective population size (*N*e) for genetic conservation and monitoring of forest trees. *Biological Conservation*, *273*, 109704. https://doi.org/10.1016/j.biocon.2022.109704

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, *22*(1), 25–33. https://doi.org/10.1016/j.tree.2006.08.009

Shaw, R. E., Farquharson, K. A., Bruford, M. W., Coates, D. J., Elliott, C. P., Mergeay, J., Ottewell, K. M., Segelbacher, G., Hoban, S., Hvilsom, C., Pérez-Espona, S., Ruņģis, D., Aravanopoulos, F., Bertola, L. D., Cotrim, H., Cox, K., Cubric-Curik, V., Ekblom, R., Godoy, J. A., … Grueber, C. E. (2025). Global meta-analysis shows action is needed to halt genetic diversity loss. *Nature*, *638*(8051), 704–710. https://doi.org/10.1038/s41586-024-08458-x

Stevenson, S. L., Watermeyer, K., Caggiano, G., Fulton, E. A., Ferrier, S., & Nicholson, E. (2021). Matching biodiversity indicators to policy needs. *Conservation Biology*, *35*(2), 522–532. https://doi.org/10.1111/cobi.13575

Stevenson, S. L., Watermeyer, K., Ferrier, S., Fulton, E. A., Xiao, H., & Nicholson, E. (2024). Corroboration and contradictions in global biodiversity indicators. *Biological Conservation*, *290*, 110451. https://doi.org/10.1016/j.biocon.2024.110451

Toszogyova, A., Smyčka, J., & Storch, D. (2024). Mathematical biases in the calculation of the Living Planet Index lead to overestimation of vertebrate population decline. *Nature Communications*, *15*(1), 5295. https://doi.org/10.1038/s41467-024-49070-x

Visconti, P., Bakkenes, M., Baisero, D., Brooks, T., Butchart, S. H. M., Joppa, L., Alkemade, R., Di Marco, M., Santini, L., Hoffmann, M., Maiorano, L., Pressey, R. L., Arponen, A., Boitani, L., Reside, A. E., van Vuuren, D. P., & Rondinini, C. (2016). Projecting Global Biodiversity Indicators under Future Development Scenarios. *Conservation Letters*, *9*(1), 5–13. https://doi.org/10.1111/conl.12159

Wauchope, H. S., Amano, T., Sutherland, W. J., & Johnston, A. (2019). When can we trust population trends? A method for quantifying the effects of sampling interval and duration. *Methods in Ecology and Evolution*, *10*(12), 2067–2078. https://doi.org/10.1111/2041-210X.13302

White, E. R. (2019). Minimum Time Required to Detect Population Trends: The Need for Long-Term Monitoring Programs. *BioScience*, *69*(1), 40–46. https://doi.org/10.1093/biosci/biy144

Willi, Y., Kristensen, T. N., Sgrò, C. M., Weeks, A. R., Ørsted, M., & Hoffmann, A. A. (2022). Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *Proceedings of the National Academy of Sciences*, *119*(1), e2105076119. https://doi.org/10.1073/pnas.2105076119

Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, *16*(8), 446–453. https://doi.org/10.1016/S0169-5347(01)02205-4

Supplementary Material

1. Rationale for parameter selection

Initial population size. We are not aware of a globally representative, accessible dataset on population level size estimates. For example, while the Living Planet Index contains data from standardized surveys of vertebrate populations, the data structure does not reveal whether the examined entity is a partial or full census of a population (usually measured as counts from a transect or similar survey method). So, we examined the genetic indicator pilot implementation, which did gather population sizes (Mastretta-Yanes & da Silva et al. 2024). Using the Dryad repository https://datadryad.org/dataset/doi:10.5061/dryad.bk3j9kdkm, we extracted data for all populations that had Ne directly estimated from genetic data or from Nc. We found that roughly half of populations are less than or equal to Nc 100, about three fourths are less or equal than Nc 1000, and 86% of populations are less than Nc 5000. Therefore, we chose to start our "low but increasing" populations at Ne 250 and our "high but decreasing" populations at Ne 750. These choices focus on the situations that may be "hard" for the indicator, where the populations approach Ne500.

Observer error. We are not aware of a database of estimates of error rates in population census data. A publication on observer error in flying foxes suggested an average undercount rate of 14.7% (Westcott & McKeown 2004), a recent evaluation of AI-classification of aerial images of whales had error rates of 6 to 20% (Green et al 2023), while an examination of observer error in marine mammal and turtle populations suggested average undercounts of about 25% (Davis et al 2022, see Table 2). We therefore chose values of 5%, 10%, and 20%.

Population growth rate. We used two resources to benchmark population growth rates. We used Mastretta-Yanes & da Silva et al. 2024, which compiled estimates of species or population level 'decline', even in absence of census data. We extracted the "overall decline" column (n=161) which is a percentage over a period of time. To get yearly population decline rates, we divided the total decline by a number of years assumed to be the period of observation, estimated as 2023 minus the first year of the "population structure" column. We found that most yearly decline rates were less than 5%, though a long tail of values extends up to over 30%. We also benchmarked against the Living Planet Index, a database of 30,000+ time series of vertebrate population size observations. The average annual growth rate in the Living Planet Database (LPI 2022) was about 0.96 (quantiles: 25% = 0.81, 50% = 1.00, 75% = 1.20). We therefore chose growth rates of 0.95 (5% decline) and 1.05 (5% increase) as reasonable values. For stable populations starting at a size of 500 (i.e., at the Ne 500 threshold), we set the average growth rate to 1.00 (0% change), where each population's size is maintained relatively constant through time.

Time steps: We simulated 25 years as this corresponds to the approximate time until 2050, when the goals of the Global Biodiversity Framework should be evaluated and hopefully achieved.

Number of populations: We recognize that some species may have one population, and some species may have hundreds of populations. We chose 12 populations because the vast majority of species analyzed in Mastretta-Yanes and da Silva et al. 2024 had 12 or fewer populations.

2. Indicator performance with mixed trends

Simulating population recovery from conservation investment

When populations approach or fall below Ne 500, they may be subject to conservation investment to help their recovery. In the simulations described in the main text, all populations within a set followed the same annual growth rate to evaluate the indicator's response to populations crossing the Ne 500 threshold. However, a set of mixed population trends may better approximate reality, where some populations or species decline while others recover for reasons that may include successful conservation actions.

To evaluate the indicator's performance under mixed population trends, we simulated three scenarios of investment in biodiversity recovery, with greater investment assumed to result in greater recovery. Using the same parameters (Table 1), we simulated three levels of investment in biodiversity conservation: (1) low investment, where 25% of declining populations recover while the remaining 75% continue declining, (2) moderate investment, where 50% of populations recover while 50% keep declining, and (3) high investment, where 75% of declining populations recover while the remaining 25% continue declining. We simulated 1000 randomisations of these mixed trend population sets with annual monitoring, and with three levels of observer error (low, medium, and high as in the previously described simulations).

All other methodological steps follow those described in the main text.

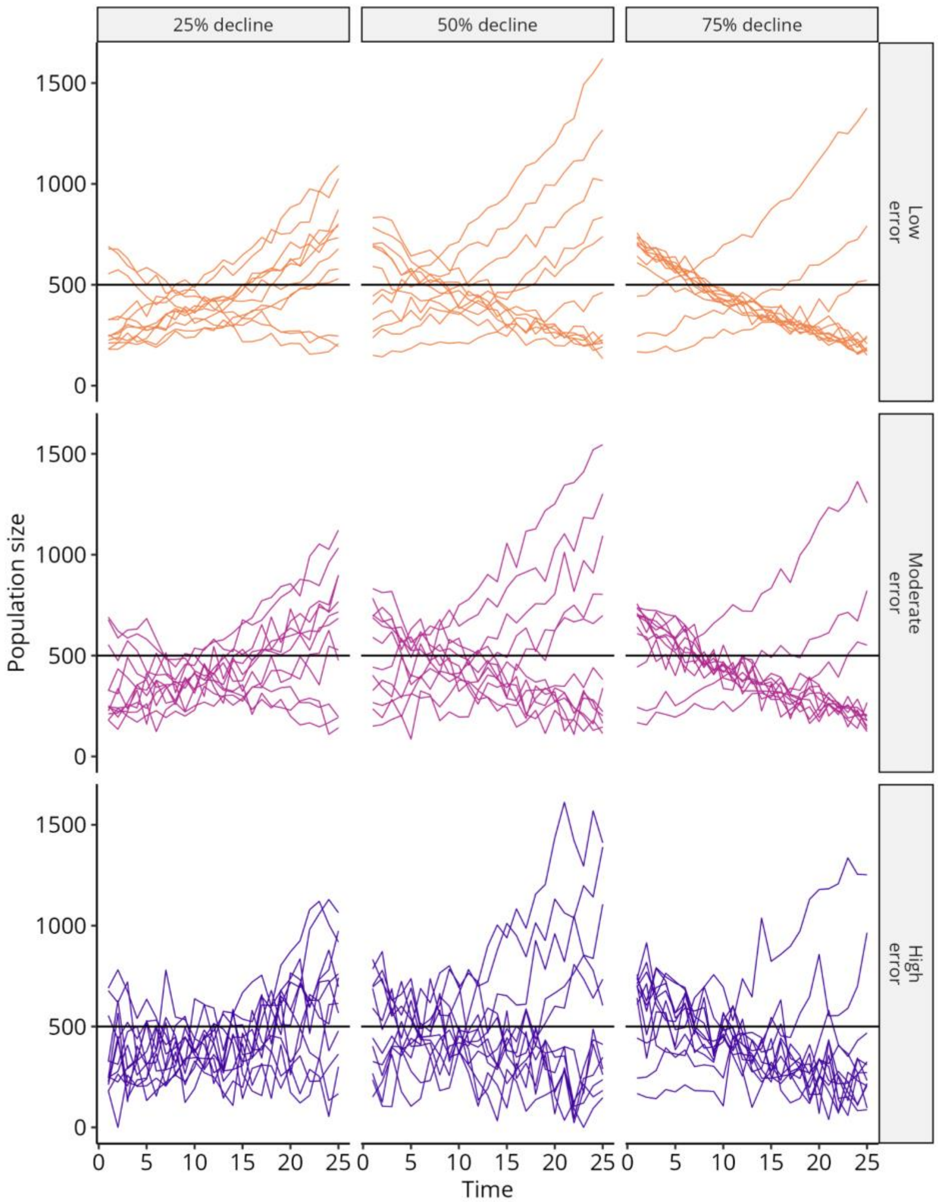


Figure S2.1. Simulations of monitored population sizes through time.

Each panel is a scenario, showing levels of conservation investment (25% recovering populations, 50% recovering populations, 75% recovering populations), observer error (Low 0.05, Moderate 0.1, and High 0.2) under annual sampling.

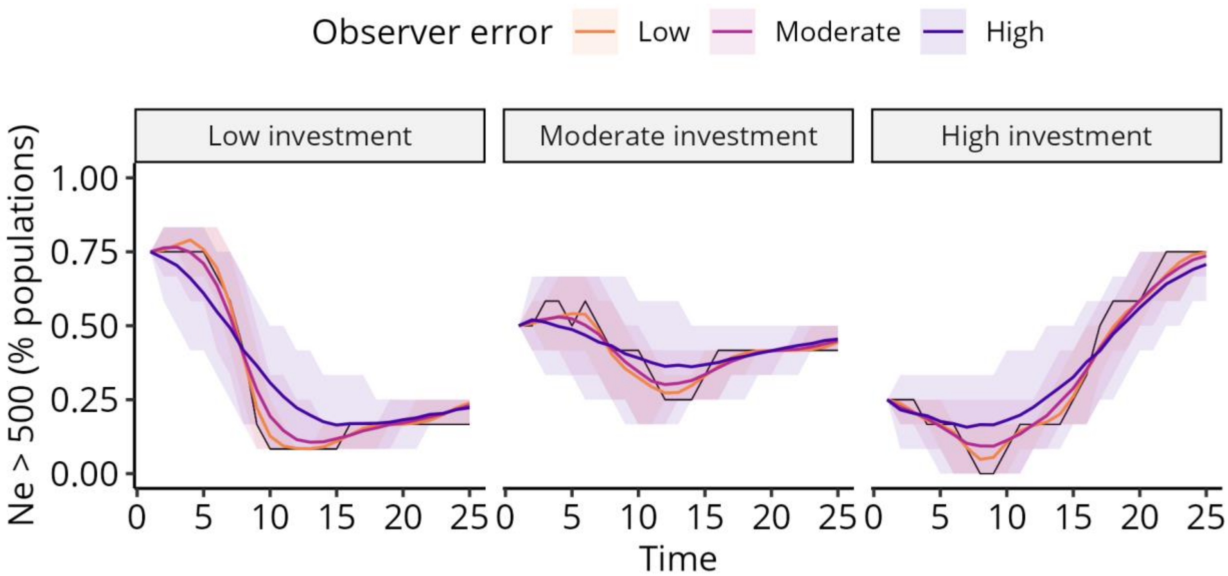


Figure S2.2. Observed Ne > 500 indicator versus the true indicator value (solid lack line) for three levels of observer error.

The true indicator value is calculated from annual data points without observer error in scenarios of conservation investment, while the observed indicator is subject to annual monitoring with observer error.

Step 1: Accurately detecting whether a population is above or below the Ne 500 threshold

When population trends were mixed (i.e., scenarios of conservation investment), we found consistently high threshold success rates regardless of the proportion of declining and growing populations, ranging from 87.3% to 97.0% success rates (Fig. 3). Threshold success rates were lowest when observer error was high (87.3% to 88.1%), and highest if observer error was low (96.7% to 97.0%).

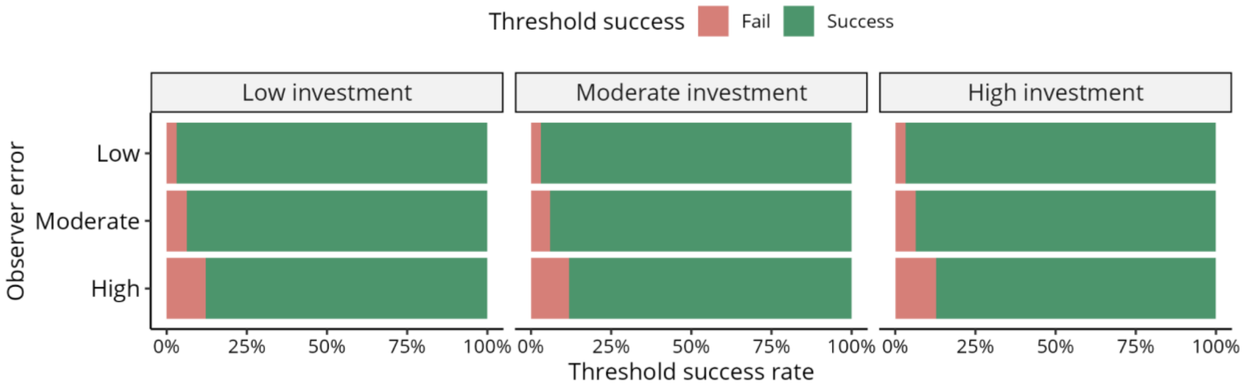


Figure S2.3. Success rate of the Ne > 500 threshold evaluation step per population over 25 years for each scenario, over 1000 randomisations. Scenarios of conservation investment with observer error, where 25% of declining populations recover (low investment), 50% recover (moderate investment), or 75% recover (high investment). For each scenario, the green bar shows the proportion of data points that were successfully classified above or below the threshold, while red shows the rate of failures during the threshold step.

Step 2: Accurately summarising population trends at the species level

When population trends were mixed under scenarios of conservation investment, accuracy remained consistently high (Fig. 4b). Accuracy generally declined with observer error as in the previous scenarios, with the best accuracy when observer error was low (0.17 ± 5.2 %) and still reasonable, but slightly lower accuracy when observer error was high (1.9 ± 11.7 %). Accuracy was the most variable, though still reasonable, when conservation investment was high (i.e., 75% of the simulated populations increased across the Ne 500 threshold) and observer error was high (2.2 ± 11.7 %).

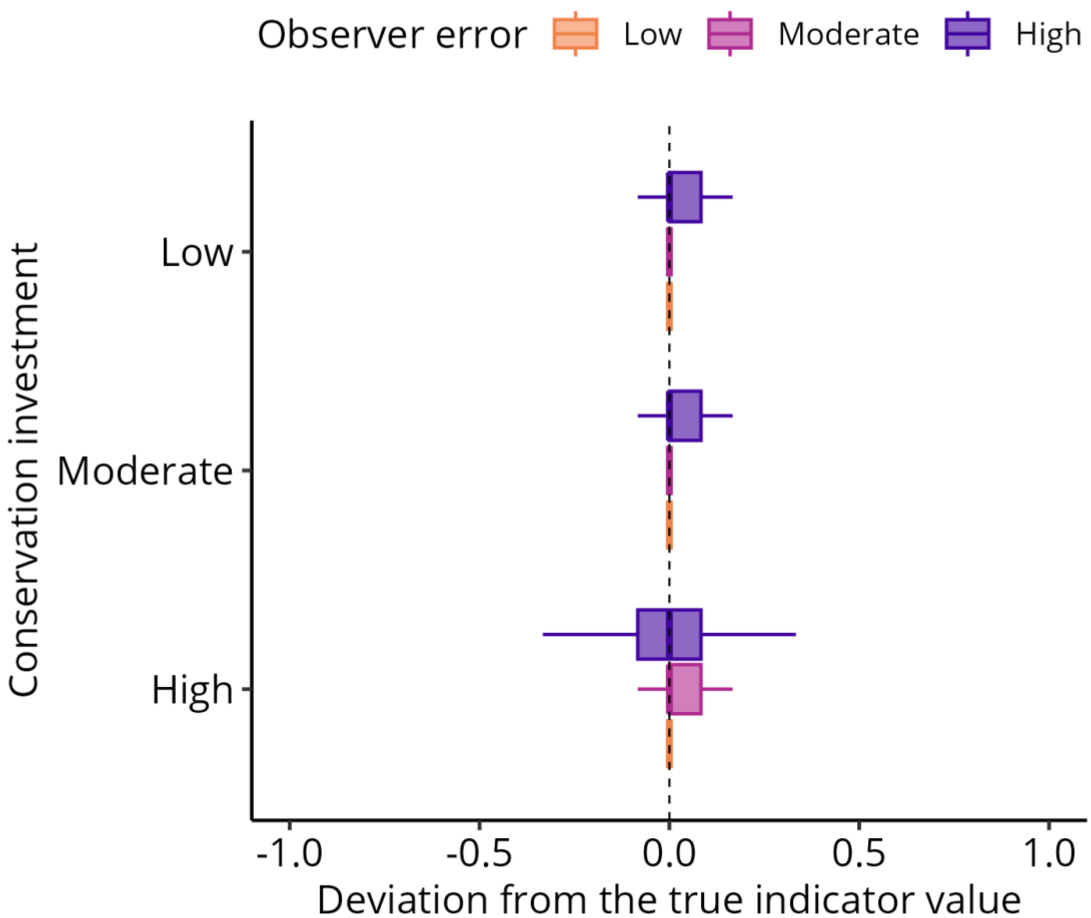


Figure S2.4. The species-level “Proportion of populations with an Ne > 500” indicator’s sensitivity to observer error and monitoring scenario. Scenarios of conservation investment with observer error, where 25% of declining populations recover (low investment), 50% recover (moderate investment), or 75% recover (high investment). Boxplots of different colours show three levels of observer error introduced into the observed population sizes that were used to calculate the observed indicator values.

3. Detection power: Alternative approaches

Extended empirical dataset

To extend the dataset to a larger number of species, such as the number of vertebrate species a country might harbor, we randomly resampled species values from the original dataset (Mastretta-Yanes et al. 2024) and added noise (using the R function *jitter()* with a factor of 50%) to create an enlarged dataset of 5830 species Ne>500 indicator values which has the same distribution as the original data:

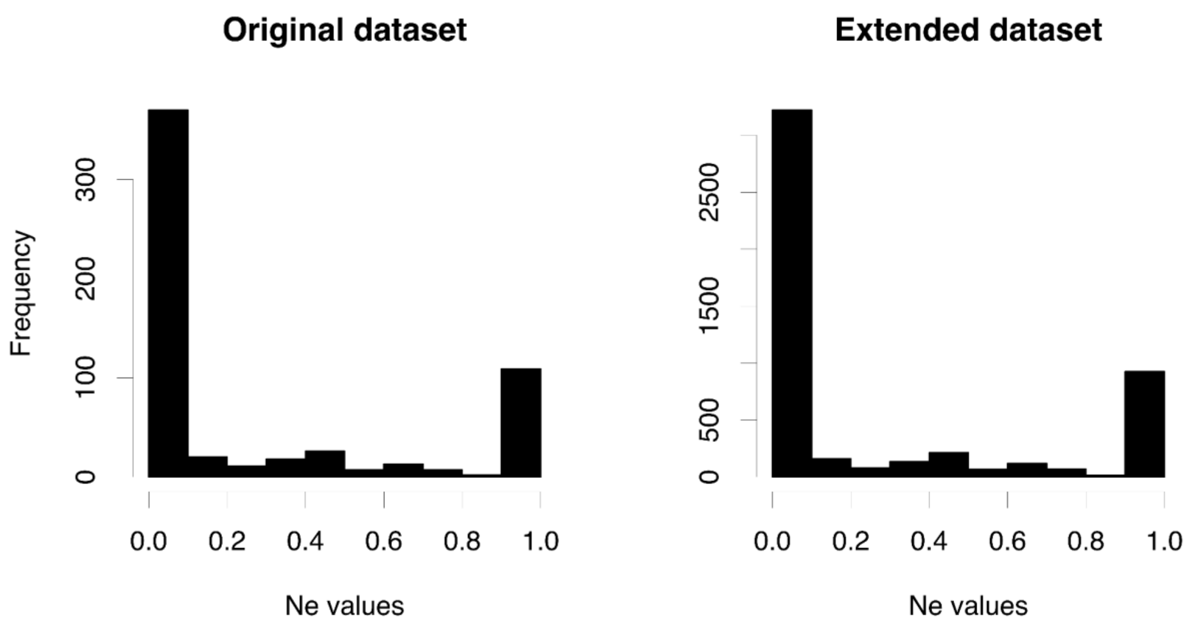


Figure S3.1. Histogram of the Ne>500 indicator values for (a) the 583 species in the Mastretta-Yanes et al. (2024) dataset and (b) the 5830 generated from the original data.

We then applied the subsampling approach described in the main text. We subsampled the extended dataset, calculated the observed country-level Ne>500 indicator for the selected species, and measured the deviance between this value to the expected value from the full dataset's country-level Ne>500 indicator (i.e., all species included).

We found that, as with a smaller dataset, the largest gains in accuracy can be made by sampling roughly 25% of the targeted species richness for species-rich regions or countries (i.e. over ~500-600 species) (See Fig. S6a). When species richness is higher than 1000, only 14% of the species need to be monitored to achieve 2.5% risk of error, and only roughly 7% of species need to be monitored to achieve 5% risk of error. Below 500-600 species, more specific recommendations can be found in the main text.

Simulated species indicator values

We applied a similar subsampling approach to simulated species-level indicator values, to evaluate the minimum number of species to monitor for an "acceptably accurate" country-level indicator under scenarios of observer error and monitoring frequency. We simulated countries with species richness ranging from 100 to 5000 (25 countries, at intervals of 200 species). Each species was composed of 12 populations with parameters from Table 1, except starting population size (N0) which was pulled from a uniform distribution between 250 and 750 individuals, and growth rate (r) which was pulled from a Normal distribution (mean = 1, SD = 0.2). For each country, we evaluated the difference between the expected "true" and the "observed" country-level indicator value for 100 randomisations of each subsample, which ranged from 10% to 100% of the country's species-level indicator values (Fig. S7).

Supplementary Tables

Table S1. Success rate of the threshold step.

Monitoring scenario	Growth rate	Observer error	Threshold outcome	Rate
Long-term	0.95	0.05	Fail	0.145
Long-term	0.95	0.05	Success	0.855
Long-term	0.95	0.1	Fail	0.156
Long-term	0.95	0.1	Success	0.844
Long-term	0.95	0.2	Fail	0.171
Long-term	0.95	0.2	Success	0.829
Long-term	1	0.05	Fail	0.358
Long-term	1	0.05	Success	0.642
Long-term	1	0.1	Fail	0.410
Long-term	1	0.1	Success	0.590
Long-term	1	0.2	Fail	0.443
Long-term	1	0.2	Success	0.557
Long-term	1.05	0.05	Fail	0.077
Long-term	1.05	0.05	Success	0.923
Long-term	1.05	0.1	Fail	0.117
Long-term	1.05	0.1	Success	0.883
Long-term	1.05	0.2	Fail	0.177
Long-term	1.05	0.2	Success	0.823
Mid-term	0.95	0.05	Fail	0.108
Mid-term	0.95	0.05	Success	0.892
Mid-term	0.95	0.1	Fail	0.118
Mid-term	0.95	0.1	Success	0.882
Mid-term	0.95	0.2	Fail	0.154
Mid-term	0.95	0.2	Success	0.846
Mid-term	1	0.05	Fail	0.344
Mid-term	1	0.05	Success	0.656
Mid-term	1	0.1	Fail	0.404
Mid-term	1	0.1	Success	0.596

Mid-term	1	0.2	Fail	0.437
Mid-term	1	0.2	Success	0.563
Mid-term	1.05	0.05	Fail	0.069
Mid-term	1.05	0.05	Success	0.931
Mid-term	1.05	0.1	Fail	0.102
Mid-term	1.05	0.1	Success	0.898
Mid-term	1.05	0.2	Fail	0.161
Mid-term	1.05	0.2	Success	0.839
Mixed	0.95	0.05	Fail	0.070
Mixed	0.95	0.05	Success	0.930
Mixed	0.95	0.1	Fail	0.091
Mixed	0.95	0.1	Success	0.909
Mixed	0.95	0.2	Fail	0.140
Mixed	0.95	0.2	Success	0.860
Mixed	1	0.05	Fail	0.327
Mixed	1	0.05	Success	0.673
Mixed	1	0.1	Fail	0.390
Mixed	1	0.1	Success	0.610
Mixed	1	0.2	Fail	0.434
Mixed	1	0.2	Success	0.566
Mixed	1.05	0.05	Fail	0.053
Mixed	1.05	0.05	Success	0.947
Mixed	1.05	0.1	Fail	0.085
Mixed	1.05	0.1	Success	0.915
Mixed	1.05	0.2	Fail	0.149
Mixed	1.05	0.2	Success	0.851
Short-term	0.95	0.05	Fail	0.035
Short-term	0.95	0.05	Success	0.965
Short-term	0.95	0.1	Fail	0.065
Short-term	0.95	0.1	Success	0.935
Short-term	0.95	0.2	Fail	0.128

Short-term	0.95	0.2	Success	0.872
Short-term	1	0.05	Fail	0.311
Short-term	1	0.05	Success	0.689
Short-term	1	0.1	Fail	0.385
Short-term	1	0.1	Success	0.615
Short-term	1	0.2	Fail	0.431
Short-term	1	0.2	Success	0.569
Short-term	1.05	0.05	Fail	0.036
Short-term	1.05	0.05	Success	0.964
Short-term	1.05	0.1	Fail	0.068
Short-term	1.05	0.1	Success	0.932
Short-term	1.05	0.2	Fail	0.135
Short-term	1.05	0.2	Success	0.865
Annual	0.95	0.05	Fail	0.031
Annual	0.95	0.05	Success	0.969
Annual	0.95	0.1	Fail	0.062
Annual	0.95	0.1	Success	0.938
Annual	0.95	0.2	Fail	0.127
Annual	0.95	0.2	Success	0.873
Annual	1	0.05	Fail	0.305
Annual	1	0.05	Success	0.695
Annual	1	0.1	Fail	0.384
Annual	1	0.1	Success	0.616
Annual	1	0.2	Fail	0.430
Annual	1	0.2	Success	0.570
Annual	1.05	0.05	Fail	0.033
Annual	1.05	0.05	Success	0.967
Annual	1.05	0.1	Fail	0.066
Annual	1.05	0.1	Success	0.934
Annual	1.05	0.2	Fail	0.134
Annual	1.05	0.2	Success	0.866

Supplementary Figures

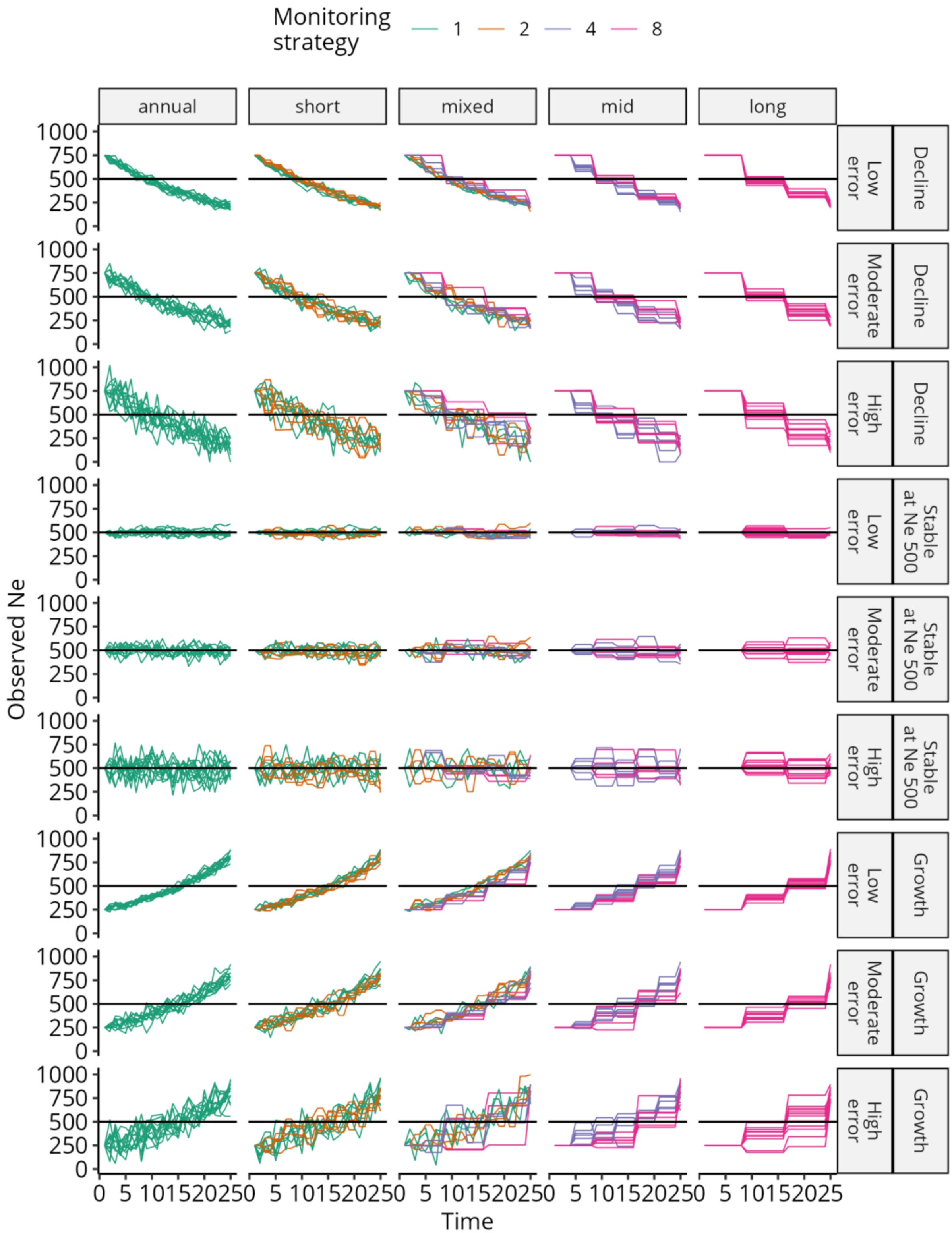


Figure S1. Simulations of monitored population sizes through time. Each panel is a scenario, showing combinations of trend direction (growth rates: decline 0.95, stable 1.00, growth 1.05), observer error (low 0.05, moderate 0.1, and high 0.2), and monitoring scenario (annual, short, mixed, mid, and long). In each scenario (panel), all populations have the same starting size and same mean growth rate with a constant variance, with one level of observer error that is applied when sampling the populations.

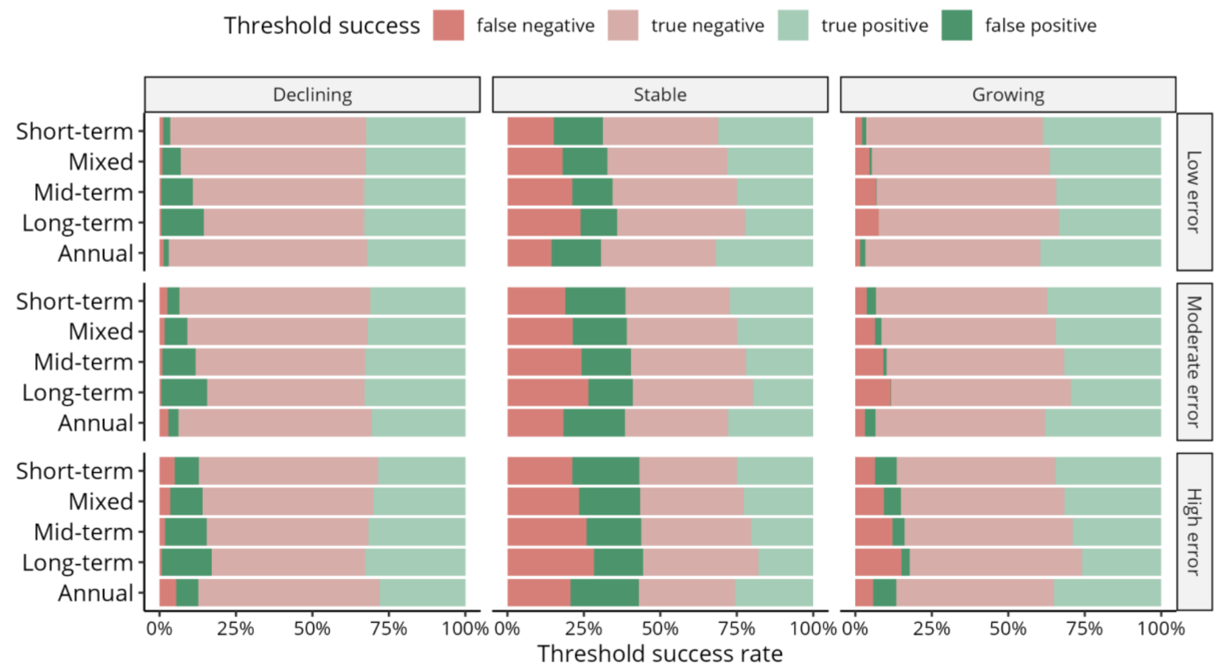


Figure S2. Success rate of classifying simulated populations as above or below the Ne > 500 threshold for all time steps and all randomizations in each scenario. Three levels of observer error were introduced to the observed population abundance time series: low (0.05), moderate (0.1), and high (0.2). False negatives mean the population is falsely classified as Ne < 500 although its true population size is above Ne 500. False positives mean the population is falsely classified as Ne > 500 although it is below Ne 500. True positives and negatives mean the population is correctly classified above or below the Ne 500 threshold, respectively.

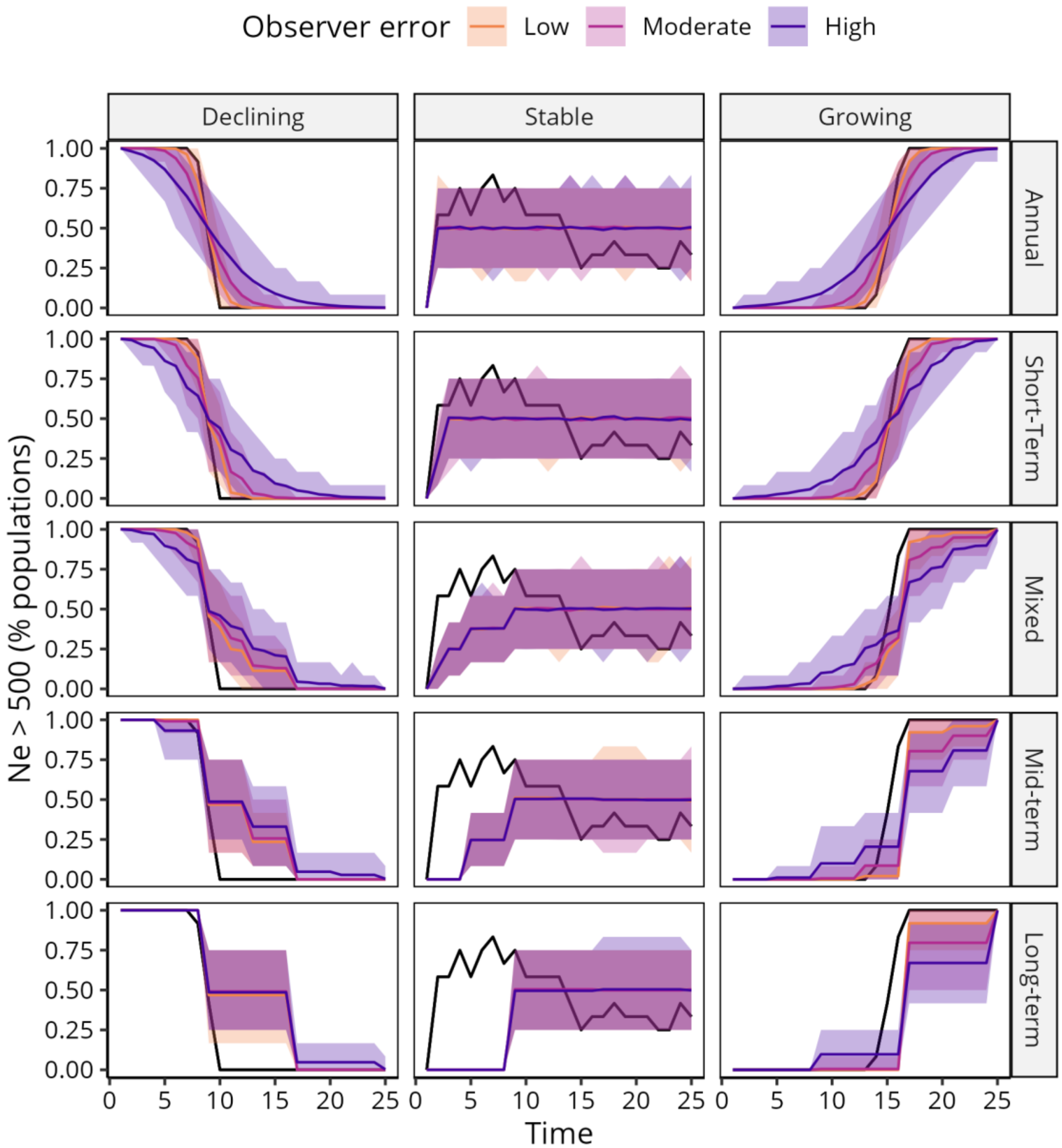


Figure S3. Expected (true) versus observed species-level Ne>500 indicator trend. The expected, or true, indicator trend is the true value based on simulated population sizes (annual data points, with no observer error). The observed indicator trend is based on simulated population sizes, measured by a simulated monitoring scheme that includes levels of observer error (colors) and different frequencies of monitoring (rows). Ribbons show the 95% confidence limits around the deviation for each scenario across all randomisations.

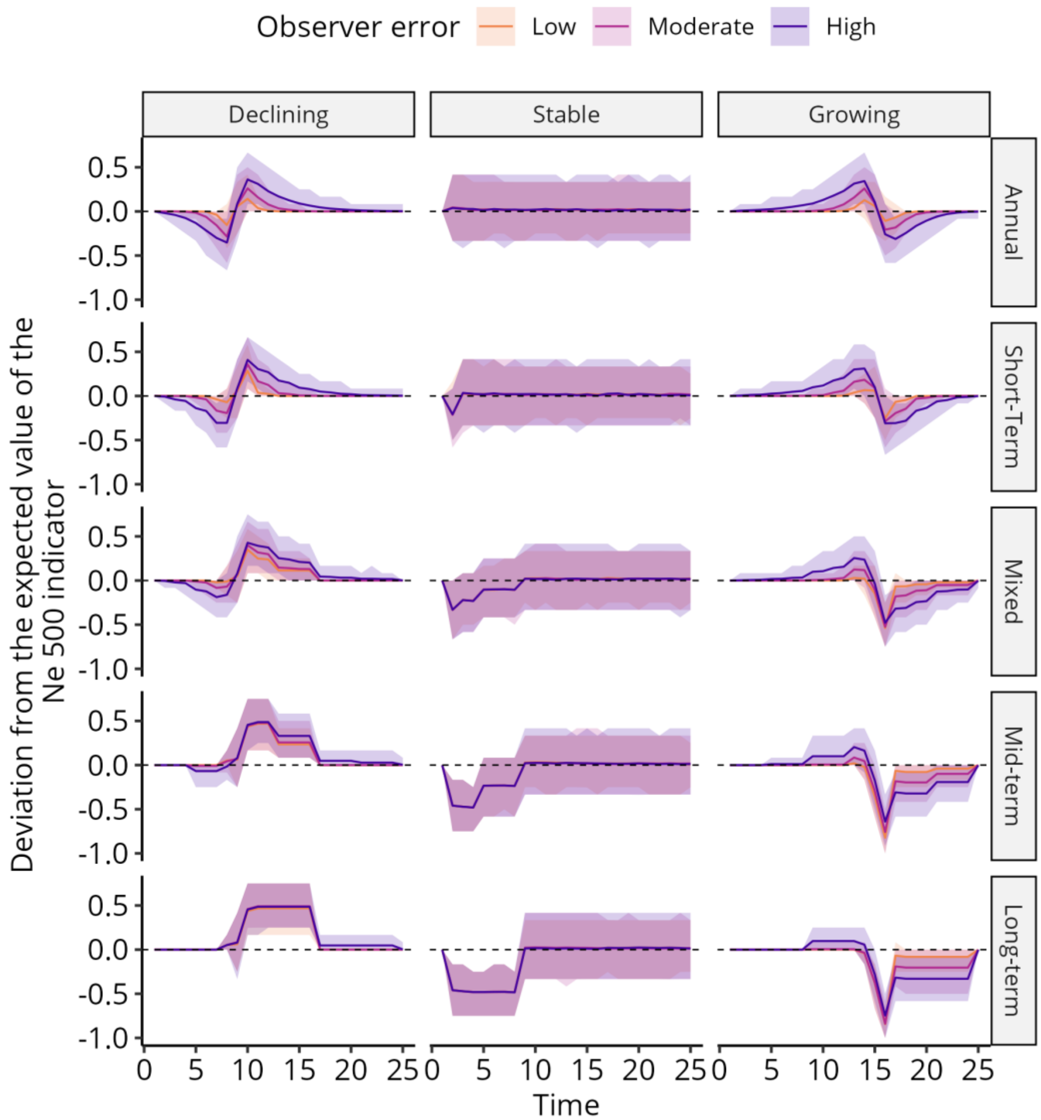


Figure S4. The indicator's deviation from the expected value in each scenario through time, over all randomisations. The deviation is the difference between the true and the expected indicator value. Ribbons show the 95% confidence limits around the deviation for each scenario across all randomisations. Negative deviation indicates that the indicator underestimates the expected value, while positive deviation indicators that the indicator overestimates the expected value.

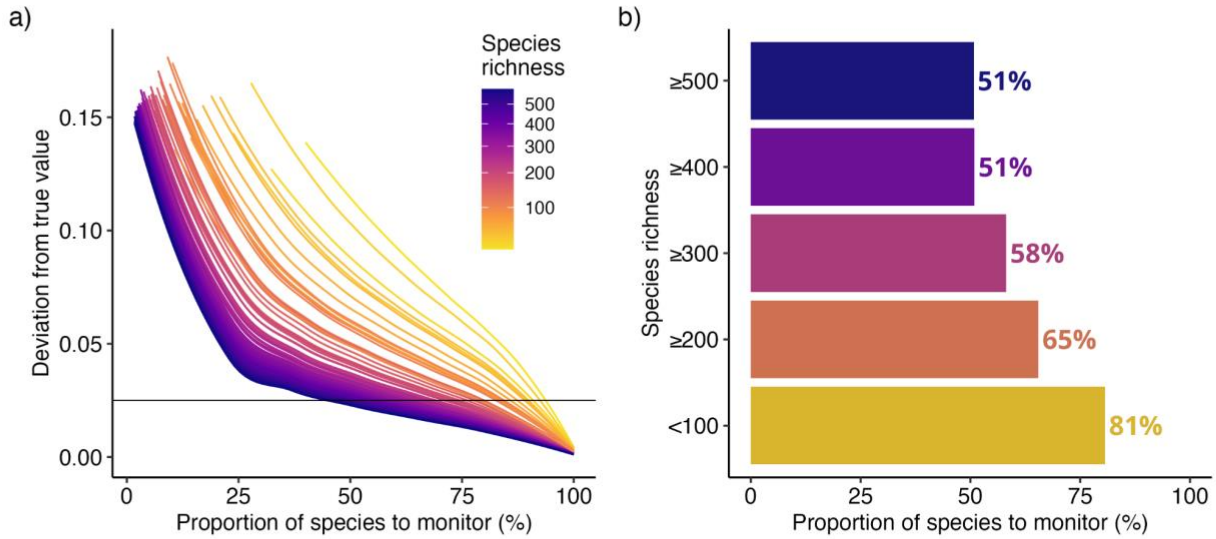


Figure S5. Accuracy of a country-level Ne>500 indicator depending on the proportion of a country's full species list that are monitored (2.5% error level). These results are based on a subsampling approach with data from Mastretta-Yanes et al. (2024). (a) Deviation between the estimated value (the indicator for a subsample of species) and the true value (the indicator for the full species list) for a range of species richness levels between 10 and 583 species, each for 10000 randomisations. We show the upper confidence limit (mean error + standard deviation) because this is a more conservative estimation of the risk of error. The horizontal line shows the proportion of species that must be monitored to risk 2.5% or less error from the true country-level indicator value, summarised in panel (b). (b) Proportion of species to monitor to risk 2.5% or less error of the country-level Ne>500 indicator as shown by the horizontal line in panel (a).

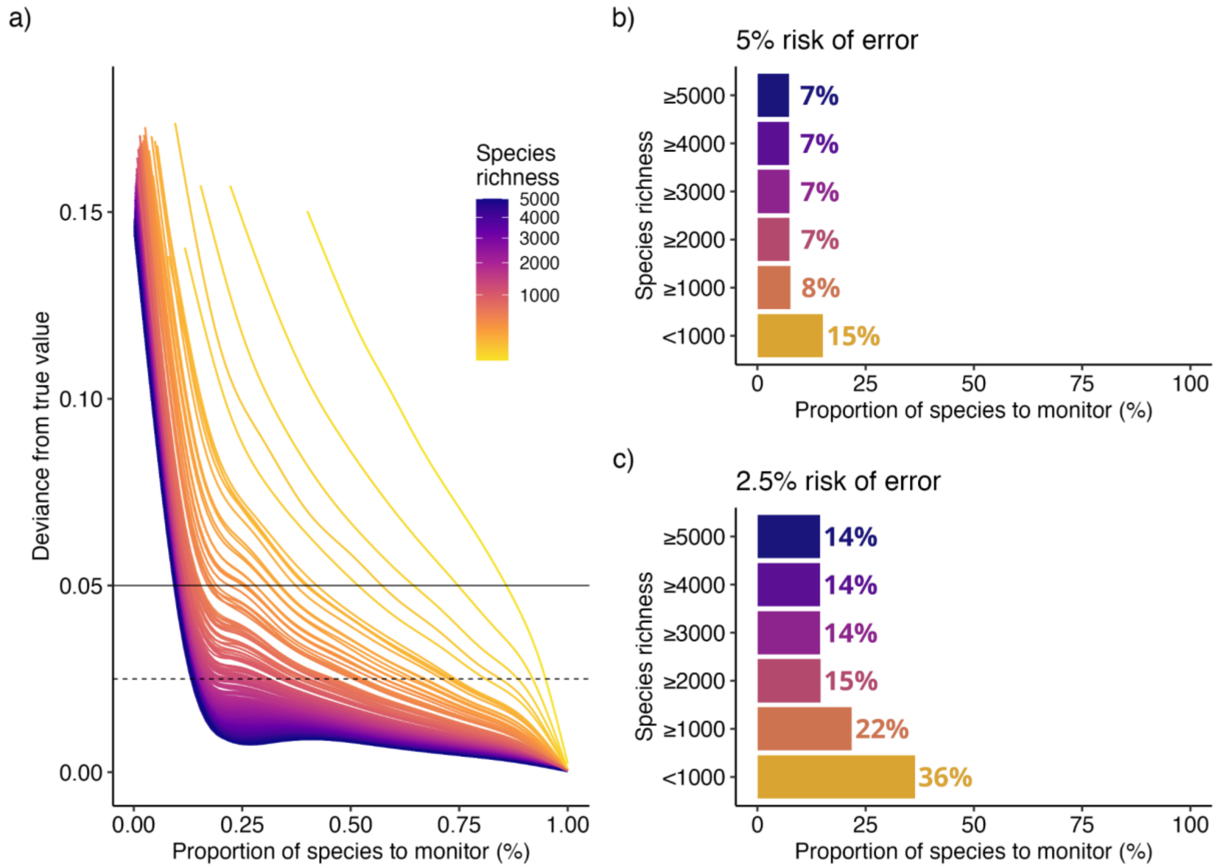


Figure S6. Accuracy of a country-level Ne>500 indicator depending on the proportion of the targeted species richness that is monitored, based on an extended dataset from Mastretta-Yanes et al. (2024) as explained in Supp. Mat. Section 3. (a) Deviation between the estimated value (the indicator for a subsample of species) and the true value (the indicator for the full species list) for a range of species richness levels between 25 and 5830 species, each for 10000 randomisations. We show the upper confidence limit (mean error + standard deviation) because this is a conservative estimation of the risk of error. The horizontal line shows the proportion of species that must be monitored to risk 5% or less (solid line) or 2.5% or less (dashed line) error from the true country-level indicator value, summarised in panels (b) and (c). Side panels show the proportion of species to monitor to risk (b) 5% or less error or (c) 2.5% or less error of the country-level Ne>500 indicator as shown by the solid and dashed horizontal lines in panel (a).

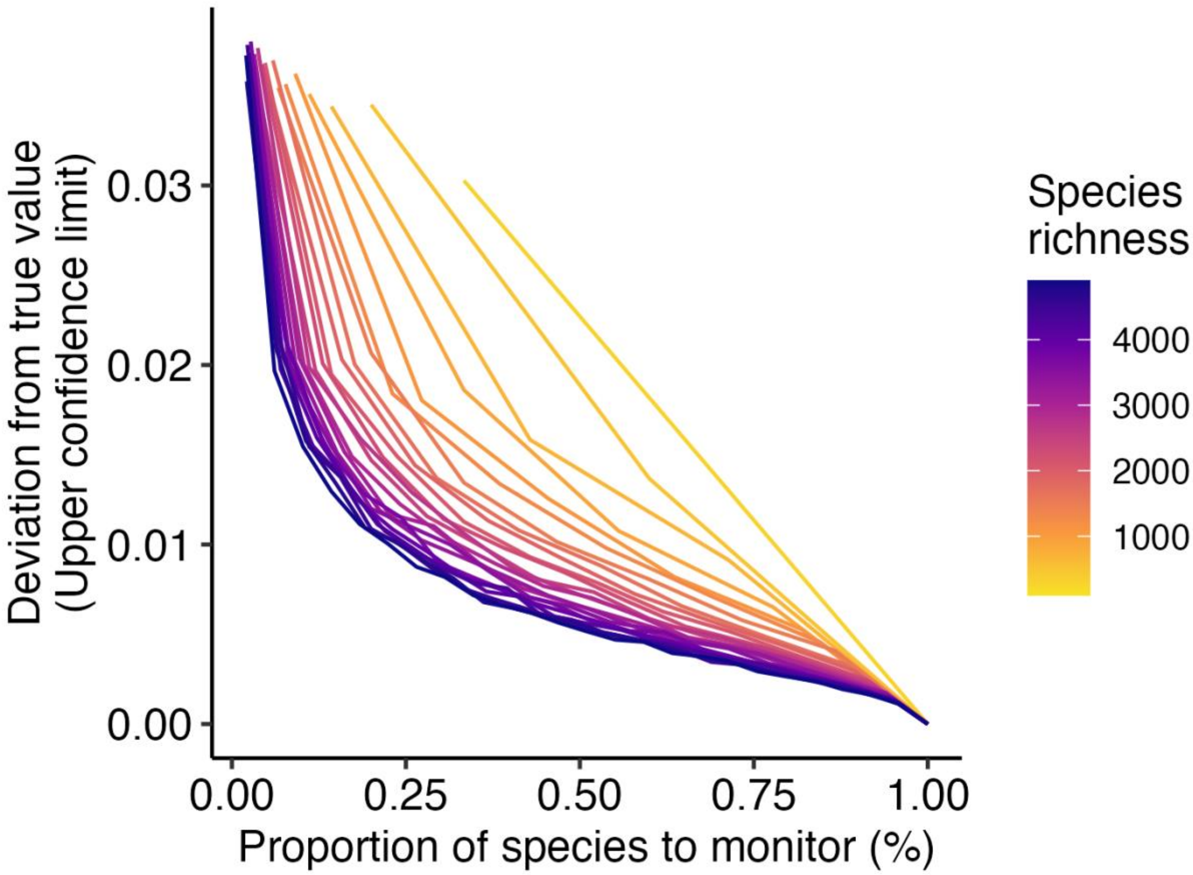


Figure S7. Accuracy of a country-level indicator based on subsampling of simulated populations. We show the upper confidence limit because this is a more conservative estimation of the relationship between sampling completeness (proportion of a country's species monitored) and indicator accuracy. Lines are coloured according to the species richness of a simulated country. Each line shows the upper confidence limit (97.5%) of the deviation between the expected "true" and "observed" indicator values from 1000 randomisations.

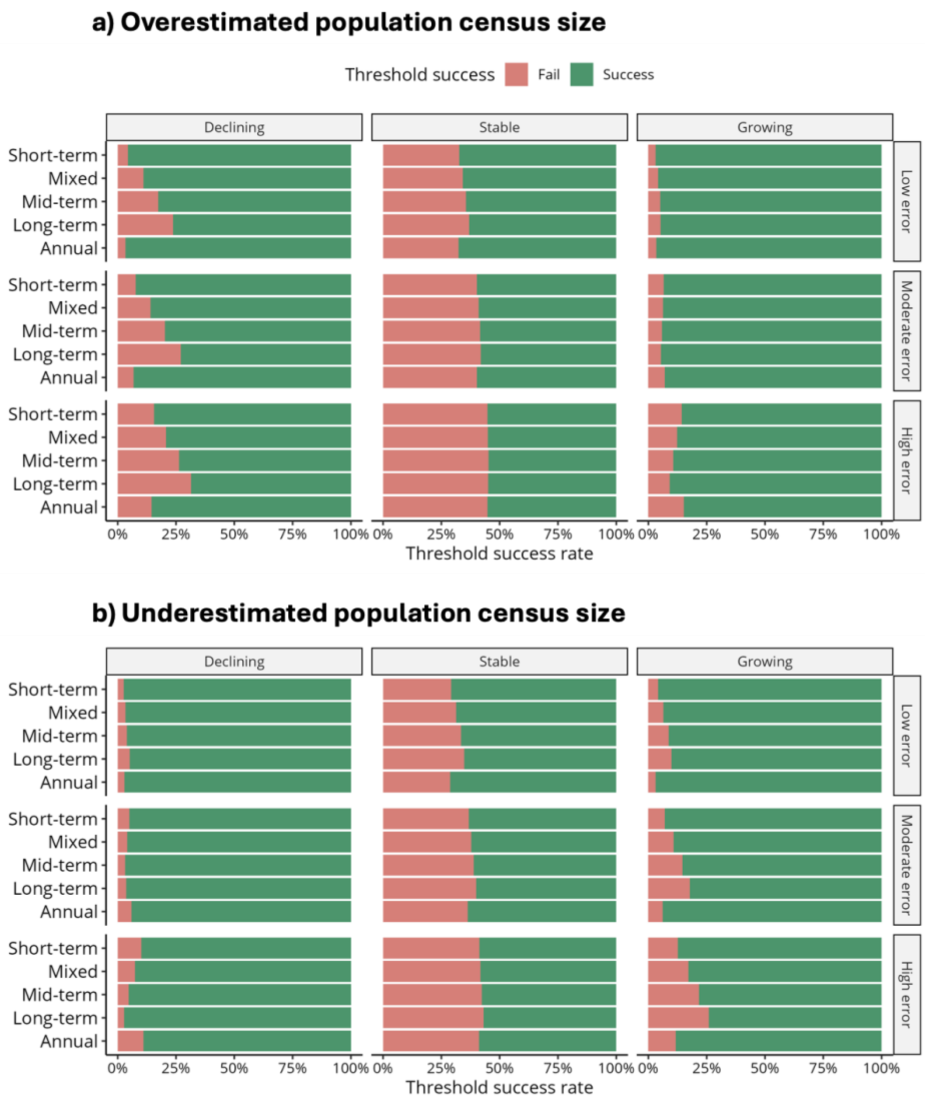


Figure S8. Success rate of the Ne > 500 threshold evaluation step in each scenario with 1000 randomisations when population census size is systematically overestimated (panel a) or underestimated (b) in the threshold. In these scenarios, the 12 populations of a species all follow the same average growth rate (declining, stable at Ne 500, growing). The green bar shows the success rate, i.e. the proportion of data points that were successfully classified above or below the threshold, while red shows the rate of failure in each scenario.

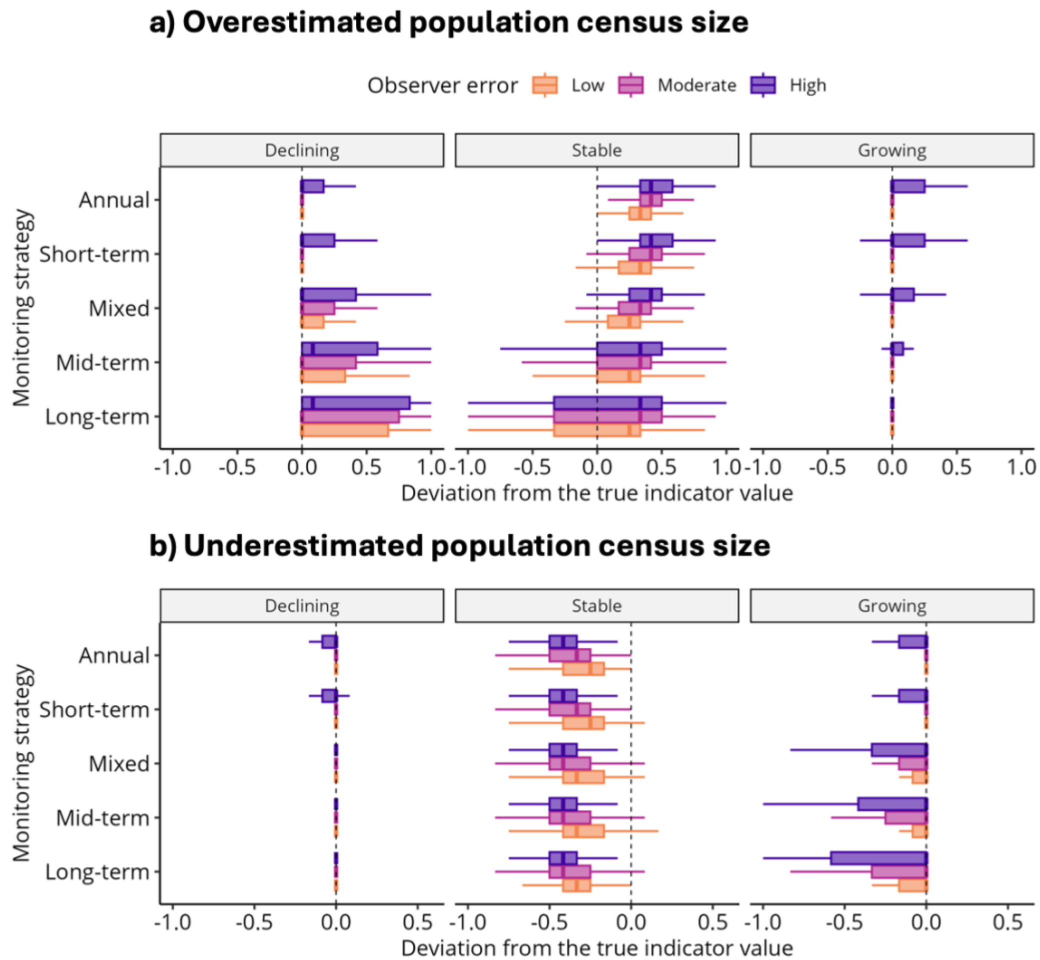


Figure S9. The species-level "Proportion of populations with an Ne > 500" indicator's sensitivity to observer error and monitoring scenario when population census size is systematically overestimated (panel a) or underestimated (panel b). Boxplots show the observed indicator's deviation from the true indicator trend in each scenario, over 1000 randomisations. Each panel shows a biodiversity change scenario in which the 12 populations of a species vary at the same annual growth rate (declining, stable at Ne 500, growing). Colours show three levels of observer error introduced into the observed population sizes that were used to calculate the observed indicator value.

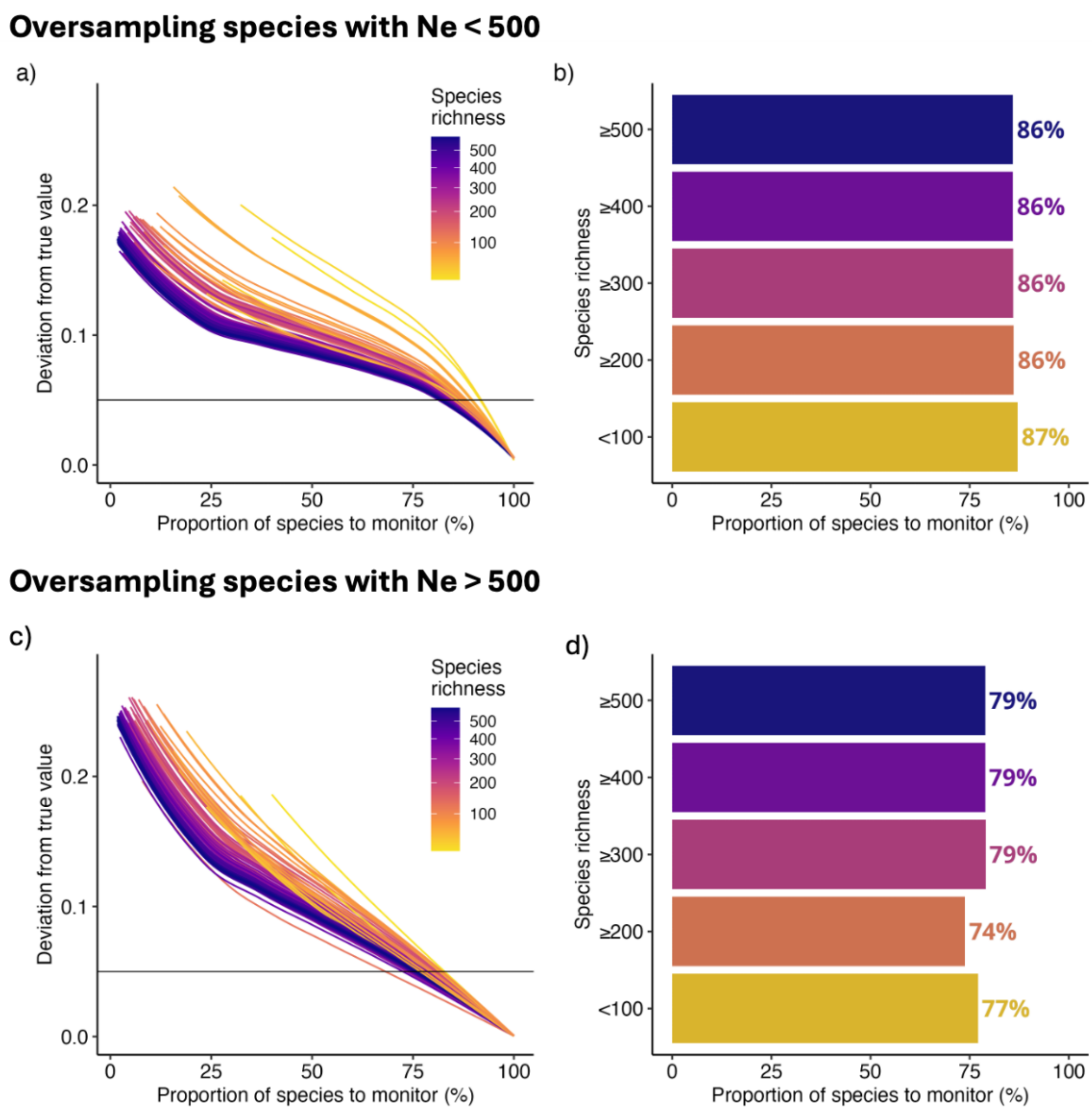

Oversampling species with Ne < 500
a)
Species richness
500
400
300
200
100
Deviation from true value
0.0
0.1
0.2
0
25
50
75
100
Proportion of species to monitor (%)
b)
Species richness
≥500
≥400
≥300
≥200
<100
86%
86%
86%
86%
87%
0
25
50
75
100
Proportion of species to monitor (%)
Oversampling species with Ne > 500
c)
Species richness
500
400
300
200
100
Deviation from true value
0.0
0.1
0.2
0
25
50
75
100
Proportion of species to monitor (%)
d)
Species richness
≥500
≥400
≥300
≥200
<100
79%
79%
79%
74%
77%
0
25
50
75
100
Proportion of species to monitor (%)

Figure S10. Accuracy of a country-level Ne>500 indicator depending on the proportion of a country's full species list that is monitored, with biased sampling of species. These results are based on a subsampling approach with data from Mastretta-Yanes et al. (2024). In panels a and b, species with Ne < 500 are over-represented (2x more likely to be monitored) in the subsample. In panels c and d, species with Ne > 500 are over-represented (2x more likely to be monitored) in the subsample. (a) Deviation between the estimated value (the indicator for a subsample of species) and the true value (the indicator for the full species list) for a range of species richness levels between 10 and 583 species, each for 10000 randomisations. We show the upper confidence limit (mean error + standard deviation) because this is a more conservative estimation of the risk of error. The horizontal line shows the proportion of species that must be monitored to risk 5% or less error from the true country-level indicator value, summarised in panel (b). (b) Proportion of species to monitor to risk 5% or less error of the country-level Ne>500 indicator as shown by the horizontal line in panel (a).

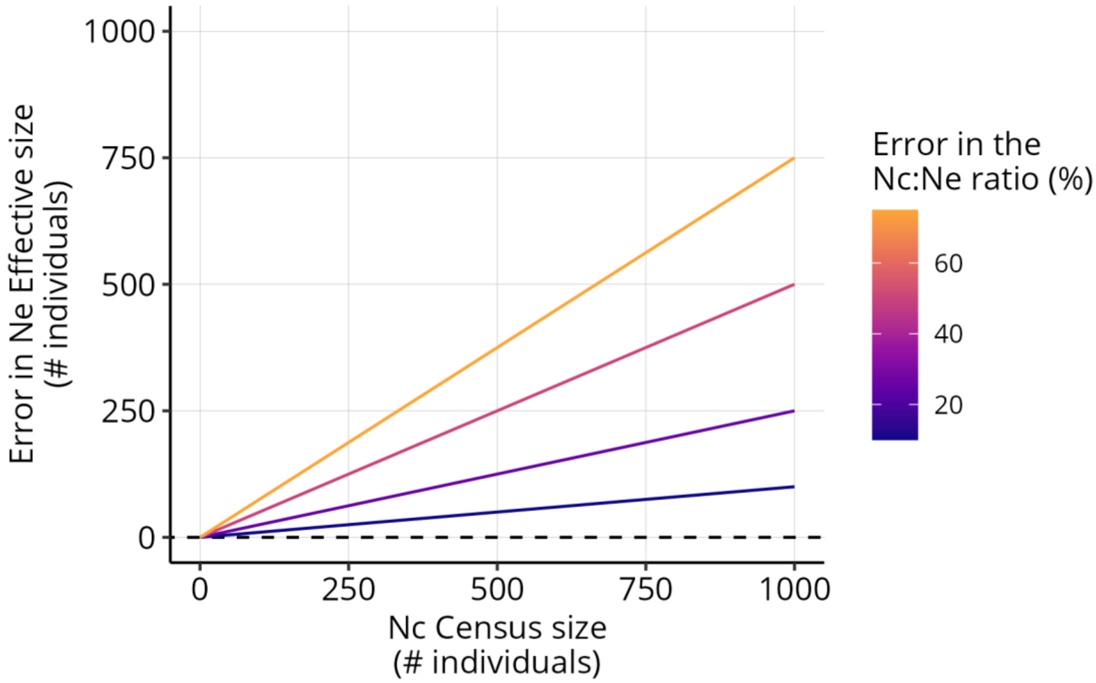


Figure S11. Demonstration of the impact of using an incorrect Ne:Nc ratio on the conversion of census size (Nc) data to effective population sizes (Ne), depending on population size. We introduced four levels of error into the Ne:Nc ratio, where the applied ratio is 10%, 25%, 50%, and 75% lower than the true ratio (which was 80%). When population census size is larger, errors in the Ne:Nc ratio have larger impacts on effective size estimation. Though we do not investigate the impacts of Ne:Nc ratio errors on the Ne>500 indicator in greater detail here, we caution that census size data should only be converted to effective population sizes when the Ne:Nc ratio is known with a high degree of confidence and accuracy to ensure that the Ne>500 indicator is a reliable metric of genetic diversity.