Title: Minimum time required to detect population trends: the need for long-term monitoring programs

Author: Easton R. White[1,2]

Address:
[1] *Center for Population Biology*
University of California, Davis
2320 Storer Hall
University of California, Davis
One Shields Avenue Davis, CA, USA

[2]Corresponding author: eawhite@ucdavis.edu , 1 480 203 7931

Number of words (including 300 words per figure): 6,950

Number of figures: 5

Number of references: 34

Supplementary material: 9 figures and 3 tables

Data and code for all the figures and tables can be found at (https://github.com/erwhite1/time-series-project).

To be submitted to: *Biological Conservation*

Available as a preprint at: https://peerj.com/preprints/3168/

Color does not need to be used for any figures in print

Abstract

Long-term time series are necessary to better understand population dynamics, assess species' conservation status, and make management decisions. However, population data are often expensive, requiring a lot of time and resources. When is a population time series long enough to address a question of interest? I determine the minimum time series length required to detect significant increases or decreases in population abundance. To address this question, I use simulation methods and examine 822 populations of vertebrate species. Here I show that on average 15.9 years of continuous monitoring are required in order to achieve a high level of statistical power. However, there is a wide distribution around this average, casting doubt on simple rules of thumb. For both simulations and the time series data, the minimum time required depends on trend strength, population variability, and temporal autocorrelation. However, there were no life-history traits (e.g. generation length) that were predictive of the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

minimum time required. These results point to the importance of sampling populations over long periods of time. I argue that statistical power needs to be considered in monitoring program design and evaluation. Short time series are likely under-powered and potentially misleading.

Keywords: ecological time series, experimental design, monitoring, power analysis, statistical power, sampling design

1 Introduction

Observational studies and population time series have become a cornerstone of modern ecological research and conservation biology (Magurran et al. 2010; Hughes et al. 2017). Long-term data are necessary to both understand population dynamics and to assess species extinction risk. Even though many time series may now be considered "long-term" (e.g. continuous plankton recorder, Giron-Nava et al. (2017)), most are still short. Time series are typically short for a variety of reasons (Field et al. 2007). They are often coupled with an experiment, which may only last a couple of years. In addition, short funding cycles make it difficult to examine populations over longer periods of time (Hughes et al. 2017).

How long of a time series is actually necessary? This question has important implications for both research and management (Nichols and Williams 2006). Scientists need to know the time series length required to address a specific question. A short time series may lead to wrong conclusions given large natural year-to-year variability (McCain, Szewczyk, and Knight 2016). Managers need to know when action is needed for a population. Therefore, managers must understand when population trend over time is actually meaningful. For example, the International Union for Conservation of Nature (IUCN) Red List Categories and Criteria suggest, under Criterion A2, a species qualifies as vulnerable if it has experienced a 30% decline over 10 years, or three generations (IUCN 2012). For both scientific and management questions, because sampling is typically expensive, we also do not want to sample for longer than is necessary. For example, Gerber, DeMaster, and Kareiva (1999) investigated the minimum time series required to estimate population growth of the endangered, but recovering, eastern North Pacific gray whale (*Eschrichtius robustus*). They used a long-term census to retroactively determine the minimum time series required to assess threat status. They found that only 11 years were needed, eight years before the delisting decision was made. This highlights the importance of estimating the minimum time series required as an earlier decision would have saved time and money (Gerber, DeMaster, and Kareiva 1999). Further, waiting too long to make a decision can imperil a species where management action could have been taken earlier (Martin et al. 2012; Martin et al. 2017).

An important step in experimental design is to determine the number of samples required. For any particular experiment four quantities are intricately linked: significance level (α), statistical power, effect size, and sample size (Legg and Nagy 2006). The exact relationship between these quantities depends on the specific statistical test. A type I error is a false positive, or incorrect rejection of a true null hypothesis. For example, if a time series was

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

assessed as significantly increasing or decreasing—when there was no true significant trend—this would be a false positive. The false positive rate, or significance level (α) is often set at 0.05 (although this is purely historical, Mapstone (1995)). A type II error (β) is a failure to detect a true trend, or failure to reject a false null hypothesis. Formally, statistical power ($1 - \beta$) is one minus the probability of a type II error (β). The effect size is a measure of the difference between two groups. Prior to an experiment, one could set appropriate levels of power, significance level, and the effect size to estimate the sample size required for the experiment. This approach, however, is not straight-forward for a time series, or more complicated scenarios (P. C. Johnson et al. 2015), as data are clearly non-independent.

For time series data, two general approaches to estimating sample size are appropriate. Simulations can be designed for a specific population and question (Bolker 2008; P. C. Johnson et al. 2015). Simple models can be simulated with parameter values corresponding to a population of interest (Gerrodette 1987). Statistical power is the proportion of simulations that meet some set of criteria. The specific criteria depend on the question at hand. For example, given a time series, when is the slope from linear regression significantly different from zero? In other words, when is the time series significantly increasing or decreasing? It is then possible to determine how power changes with a variable of interest. For example, time series can be simulated for different lengths of time. From these simulations, the minimum time series length required to meet certain levels of statistical significance and power is estimated (Bolker 2008).

In addition to using simulations, empirical time series can also be used. Multiple replicates of similar populations are usually not available, but it is possible to subsample an empirical time series (Gerber, DeMaster, and Kareiva 1999; Brashares and Sam 2005). Subsamples of different lengths can then be evaluated to estimate the proportion of subsamples meeting some criteria, again a measure of statistical power. Similar to the simulation approach, this measures of power can be used to determine the minimum time series required for a particular question of interest.

Past work has investigated questions related to the minimum time series required to estimate trends in population size over time (Wagner, Vandergoot, and Tyson 2009; Giron-Nava et al. 2017). For example, Rhodes and Jonzen (2011) examined the optimal allocation of effort between spatial and temporal replicates. Using simple populations models, they found that the allocation of effort depends on environmental variation, spatial and temporal autocorrelation, and observer error. Rueda-Cediel et al. (2015) also used a modeling approach, but parameterized a model specific for a threatened snail, *Tasmaphena lamproides*. They found that for this short-lived organism, 15 years was adequate to assess long-term trends in abundance. However, these studies, and other past work, have typically focused either on theoretical aspects of monitoring design or focused on only a few species.

I use both simulations and empirical time series to determine the minimum number of years required to address several questions. I estimate the minimum time series length required (T_{min}) to assess long-term changes in abundance via simple linear regression. First, I estimate T_{min} using a simulation approach. Then I examine 822 population time series to estimate T_{min}. In the supplementary material, I determine T_{min} for related ideas: using more complicated population models, varying statistical level and power, and the use of generalized additive

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

models.

2 Methods

2.1 Simulation approach

One approach to determining the minimum time series length needed is through repetitive simulations of a population model (Gerrodette 1987). This is the same approach one might use in sample size calculations for any experimental design too complicated for simple power analyses (Bolker 2008; P. C. Johnson et al. 2015). I only briefly discuss this approach as it has been described elsewhere. Essentially, a population model is simulated repetitively for a number of years. This approach requires us to determine values for model parameters (e.g. population variability). As an example, we can take the following population model for population size N at time t:

$$N(t+1) = N(t) + r(t) + \epsilon \text{ with } \epsilon \sim N(\mu, \sigma) \quad (1)$$

where ϵ is a normally-distributed random noise term with mean μ and standard deviation σ. The rate of growth r is also the trend strength of the increase or decrease (i.e. the rate of increase). It is important to note that any population model could be substituted for equation 1, as in the supplementary material (Figs. A6, A7).

Statistical power is then the proportion of simulations that meet some criteria. Here, our criteria is whether the slope parameter from linear regression is significant at the 0.05 threshold with statistical power of 0.8. Statistical power of 0.8 would indicate that, if there was a true trend in abundance, there would be a 0.8 probability of detecting the trend. Values of 0.05 for the significance level and 0.8 statistical power are historical and it is important to examine the effect of changing these values (Fig. A4).

In Fig. 1a, a number of simulated time series are shown for a set number of time periods ($t = 40$). It is clear that statistical power increases quickly with increases in length of time sampled (Fig. 1b). Where power is greater than 0.8 (the dotted line), that is the minimum time required (T_{min}) to be confident in the detection of a long-term trend in abundance.

2.2 Data source

I use a database of 2444 population time series compiled in (Keith et al. 2015); they compared the predictability of growth rates among populations. The data are originally from the Global Population Dynamics Database (NERC Centre for Population Biology 2010) and several other sources (Keith et al. 2015). I filtered out short time series (less than 35 years), and those with missing data, leaving 822 time series. The data includes information on 477 vertebrate species with a focus on mammals, birds, and fish. The data also includes information on

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

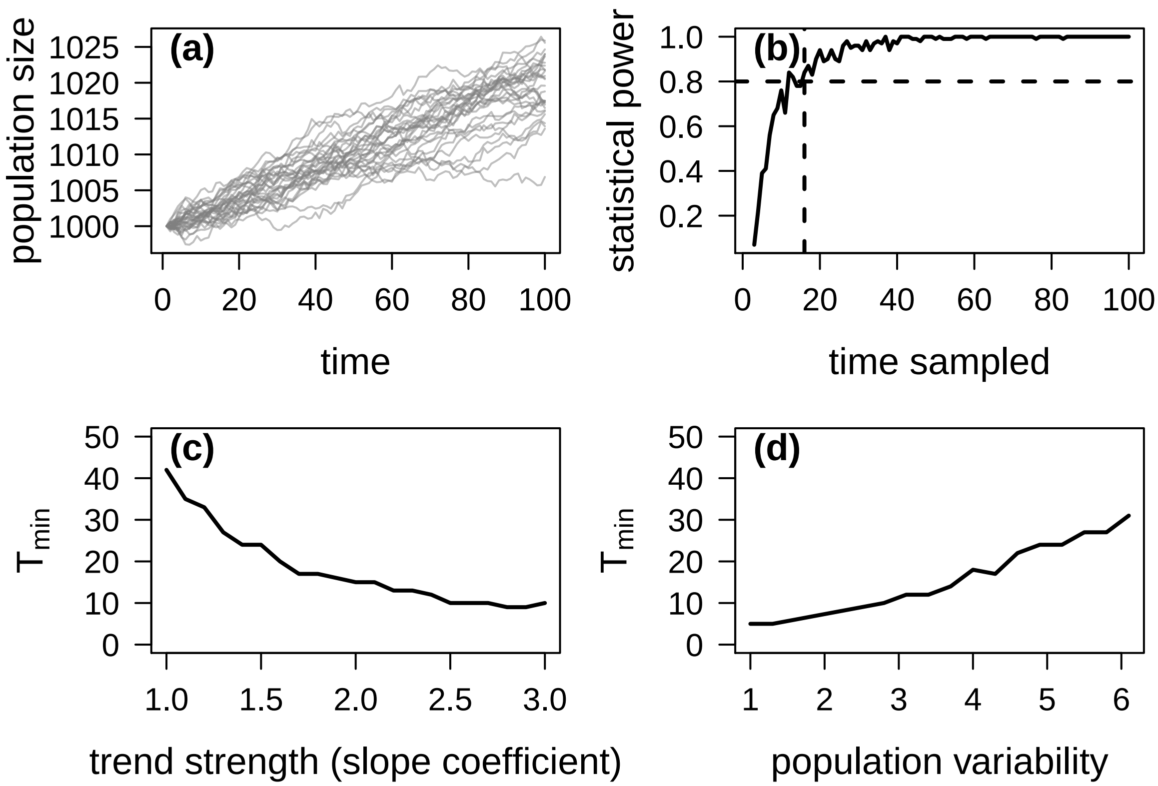

Figure 1: (a) Example of a simulated time series for 40 time periods. (b) Statistical power versus the simulated time series length. The horizontal, dashed line is the desired statistical power of 0.8. The vertical, dashed line is the minimum time required to achieve the desired statistical power. (c) Minimum time required (T_{min}) for simulations with different values of the trend strength (r). (d) Minimum time required for different levels of population variability (σ). In each case, the minimum time required is the minimum number of years to achieve 0.8 statistical power given a significance level of 0.05.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

generation length and survey specifications. For each time series, I also calculate other variables of interest: coefficient of variance in population size, long-term trend in abundance (slope coefficient from simple linear regression), and temporal autocorrelation. All analyses were conducted in R (R Core Team 2016).

For a subset of populations ($n = 547$), there is information on life-history traits available from another paper (Myhrvold et al. 2015), including body size and generation length. All 547 populations were birds. I examine how the minimum time required is related to these life-history traits (Fig. 4).

2.3 Empirical approach

I assume that each time series is long enough to include all necessary information (e.g. variability) about the population. In other words, each time series is a representative sample. I first take all possible contiguous subsamples of each time series. For example, a time series of 35 years would have 34 possible contiguous subsamples of length 2, 34 possible contiguous subsamples of length 3, and continuing until 1 possible contiguous subsample of length 35 (Gerber, DeMaster, and Kareiva 1999; Giron-Nava et al. 2017). Next, I run a linear regression for each subsampled time series. Then, I determine the proportion of subsamples of a particular length that have estimated slope coefficients which are statistically different from zero. I only look at the proportion of samples where the long-term, or "true", time series also has a significant slope. This proportion is a measure of statistical power. Lastly, I determine which subsample length is required to achieve a certain threshold of statistical power (0.8, Cohen (1992)). The minimum subsampled length that met these criteria is the minimum time series length required (T_{min}).

In the supplementary material, I show how the same approach described here for more complicated population models. I also determine the minimum time required to estimate long-term trends according to generalized additive models, instead of the simple linear models used here (Fig. A8).

3 Results

I determined the minimum time series length (T_{min}) required to address a particular question of interest. What is the minimum time series length required to determine, via linear regression, the long-term population trend? Here, the minimum time series length required had high enough statistical power (greater than 0.8) for a set significance level (α) of 0.05. It is also possible to alter statistical power and α. Predictably, with increased statistical power or decreased α, T_{min} increased (Fig. A4). I then estimated T_{min} using two approaches. I briefly describe results from the simulation approach and then discuss the empirical approach.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

3.1 Simulation approach

I constructed a general population model where the trend strength (i.e. slope coefficient) over time could be a model parameter. I then simulated time series of different lengths. From these simulations I determined the minimum time series length required to achieve a certain level of statistical power. In line with past work (Gerrodette 1987), I found the T_{min} increases (i.e. more time is required) with decreases in trend strength and with increases in population variability (Figs. 1c,d).

I chose a simple model, but any other population model could be used (see example in Fig. A6). Ideally, the specific model choice should be tailored to the population of interest. I explored how the simulation approach could be applied to more biologically-realistic population models (Fig. A7). Specifically, I determined the minimum time required to estimate long-term population trends using a stochastic, age-structured model of lemon shark population dynamics in the Bahamas (White, Nagy, and Gruber 2014). I found that over 27 years of continuous monitoring were needed in this particular scenario (Fig. A7). Similar to the simulation approach described above, the minimum time required for the lemon shark population was strongly dependent on model parameters.

3.2 Empirical approach

I examined a database of 822 separate population time series representing 477 species. This database consisted of vertebrate species with a variety of life-history characteristics (Fig. 4). I limited analyses to populations with at least 35 years of continuous sampling. I then examined the minimum time required to estimate long-term trends via linear regression.

Across all the populations, I found an average minimum time series length required (T_{min}) of 15.9 (SD=8.3), with a wide distribution (Fig. 2b). Estimates of T_{min} varied between biological class (Fig. 2a). Ray-finned fish (class Actinopterygii) typically had estimates of T_{min} over 20 years. Birds (class Aves) had a much wider distribution of T_{min}, but usually required less years of sampling. Differences between these classes can be explained by differences in variability in population size and strength of trends in abundance (Fig. A3).

3.2.1 Corrrelates for minimum time required

The minimum time series length required was strongly correlated with trend strength (i.e. estimated slope coefficient from linear regression), coefficient of variation in population size, and autocorrelation in population size (Fig. 3). This is in line based on simulations here and those of others (Rhodes and Jonzen 2011). Using a generalized linear model, with a Poisson error structure, all three of these explanatory variables were significant and had large effect sizes (see Table A1). Combined, trend strength, coefficient of variation in population, and autocorrelation account for 75.1% of the explained deviance (Zuur et al. 2009) in minimum time series length required.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

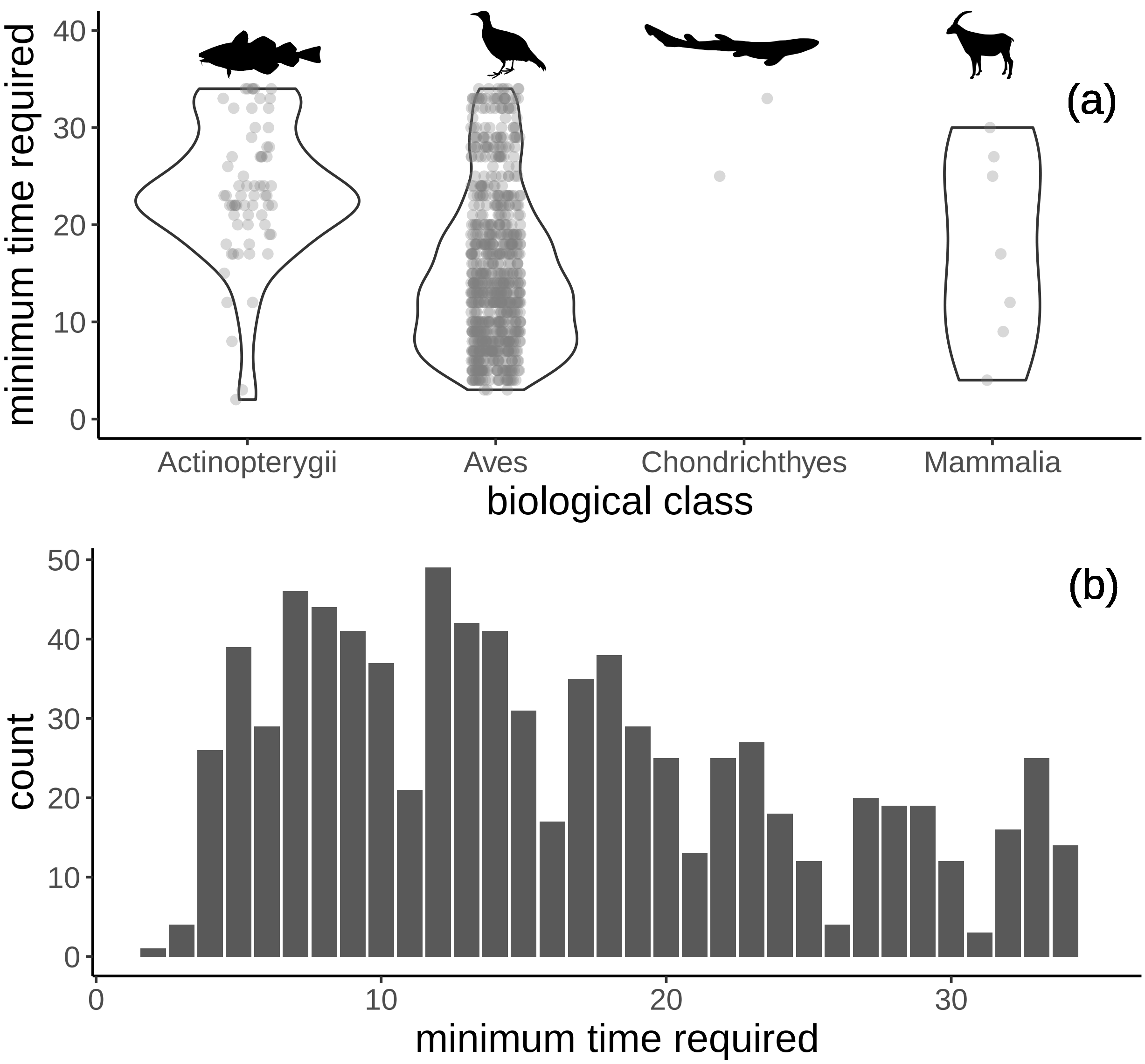

Figure 2: (a) Distributions of the minimum time required for populations from four different biological classes. (b) Distribution of minimum time required for all populations regardless of biological class. The minimum time required calculation corresponds to a significance level of 0.05 and statistical power of 0.8.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

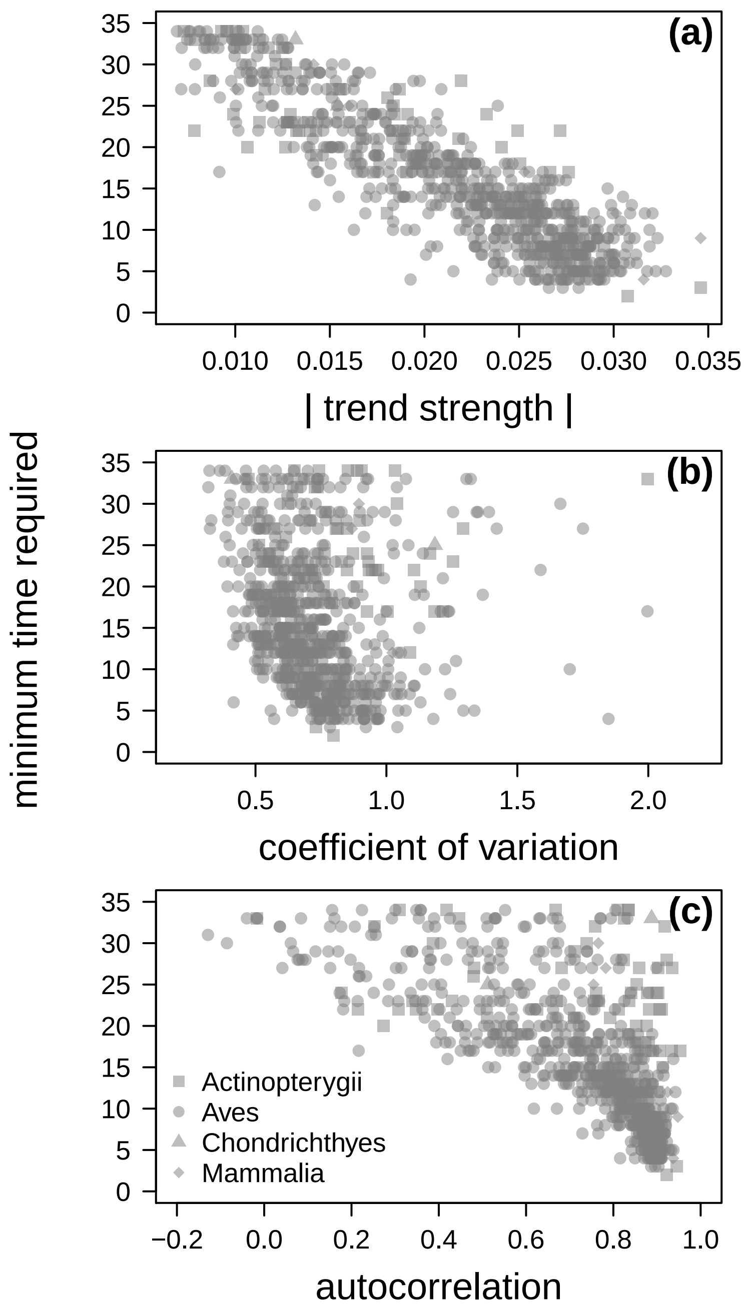

Figure 3: Minimum time required to estimate change in abundance correlated with (a) trend strength (absolute value of slope coefficient estimated from linear regression), (b) coefficient of variation in interannual population size, and (c) temporal lag-1 autocorrelation.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

For a subset of the populations I combined time series data with a data on life-history characteristics of amniotes (Myhrvold et al. 2015). There was life-history information available for 547 populations representing 315 different species, all of which were birds (Aves class).

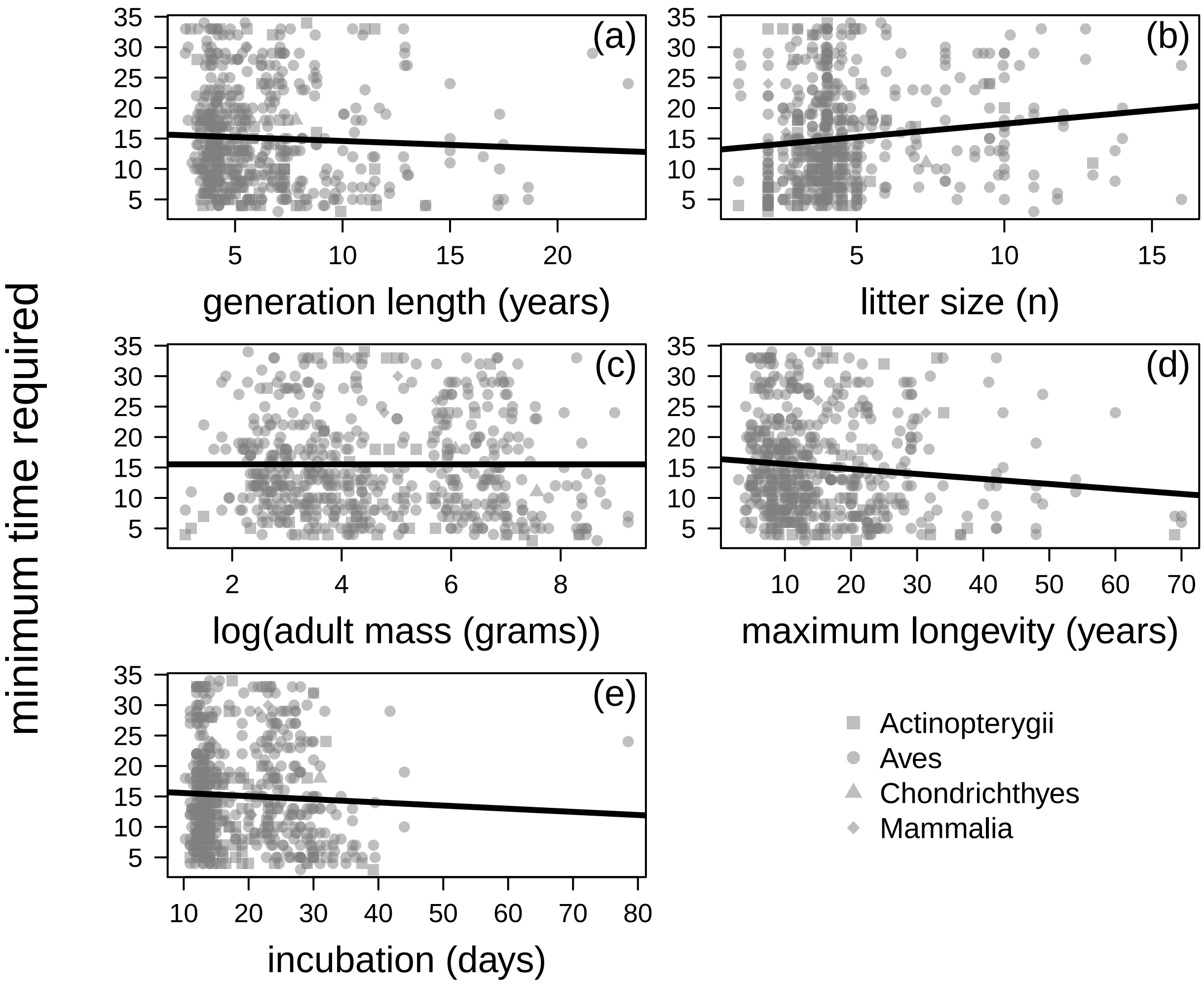

Figure 4: Mimimum time required versus (a) generation length (years), (b) litter size (n), (c) log adult body mass (grams), (d) maximum longevity (years), and (e) incubation (days). The lines in each plot represent the best fit line from linear regression.

Some life-history traits were significant predictors for the minimum time required (Fig. 4, Tables A2,A3). However, none of these life-history traits explained a large part of the variation in minimum time required. In a generalized linear model, all of the life-history traits described in figure 4 account for only 5.99% of the explained deviance in minimum time series length required. In addition, when accounting for trend strength, coefficient of variation, and autocorrelation, no life-history traits were significant predictors of the minimum time required (Table A3).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

3.2.2 Evalulating the IUCN criteria

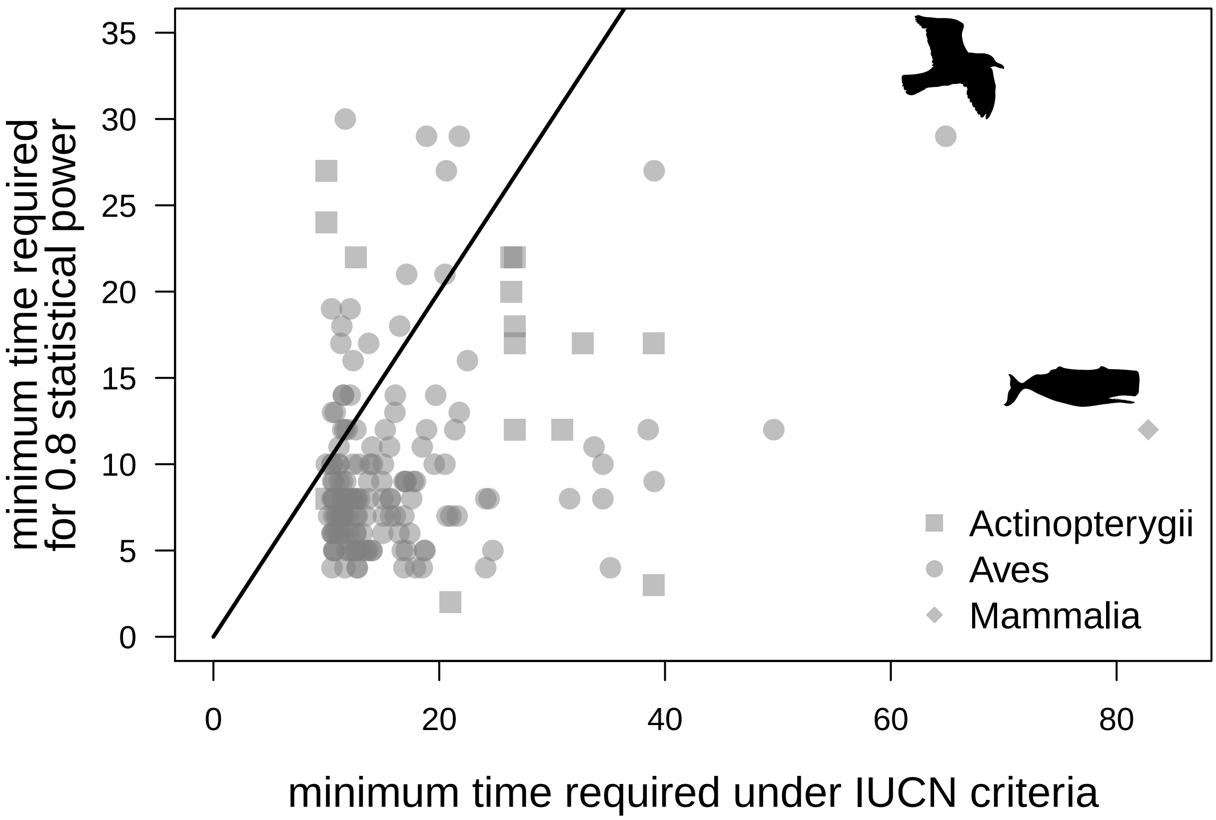

Figure 5: Minimum time required to achieve 0.8 statistical power versus the minimum time required under IUCN criteria A2 to classify a species as vulnerable. Each point represents a single population, all of which saw declines of 30% or greater over a 10 year period.

The IUCN Red List Categories and Criteria suggest, under Criterion A2, a species qualifies as vulnerable if it has experienced a 30% decline over 10 years, or 3 generations (whichever is longer) (IUCN 2012). I examined a subset of populations with observed declines of 30% or greater over 10 years, qualifying all of them as vulnerable. This resulted in n = 162 populations of fish, birds, and a single mammal. I then compared the minimum time required to achieve 0.8 statistical power (T_{min}) to the minimum time required under the IUCN criteria (Fig. 5). For populations below the identity line in figure 5, IUCN criteria would require more sampling compared to estimates for T_{min}. Further, populations above the identity line are cases where the IUCN criteria would classify a population as vulnerable despite not having sampled enough years to achieve high statistical power (Fig. 5). The silhouettes on figure 5 highlight that species with longer generation times typically have larger discrepancies between T_{min} and the minimum time required for IUCN assessments (Fig. A5).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

3.2.3 Sensitivity analysis

Lastly, I tested model sensitivity by using generalized additive models (GAMs) instead of simple linear regression. Again, I examined the minimum time required to estimate long-term population trends (Fig. A8). I found that although I obtain a similar distribution of minimum times required for GAMs, the minimum time required for GAMs is on average shorter than for linear regression (Fig. A9).

4 Discussion

I explored two approaches to estimate the minimum time series length required to address a particular question of interest. I asked, what is the minimum time series length required to determine long-term population trends using linear regression? This is one of the simplest questions one could ask of a time series. The simulation-based approach has been suggested by others, especially in situations more complicated than that suited for classic power analysis (Gerrodette 1987; P. C. Johnson et al. 2015; Bolker 2008). My simulations support past work that longer time series are needed when the trend strength (i.e. rate of increase or decrease) is weak or when population variability is high (Gerrodette 1987). I also showed how the simulation model can be altered for a particular population (Fig. A7) or question (Figs. A6,A8).

Here, I focus on an empirical approach to estimate the minimum time series length required to assess changes in abundance over time. I examined 822 population time series (all longer than 35 years). I then subsampled each to determine the minimum time required to achieve a desired significance level and power for linear regression. Statistical power is important as it provides on information as to the necessary samples required to determine a significant trend (Legg and Nagy 2006). I found that on average 15.91 years of continuous monitoring were typically necessary (Fig. 2b). However, the distribution of minimum time required was wide. This time-frame is in line with past work on a short-lived snail species (Rueda-Cediel et al. 2015) and a long-lived whale species (Gerber, DeMaster, and Kareiva 1999). Hatch (2003) used seabird monitoring data to estimate minimum sampling requirements. He found that the time required ranged from 11 to 69 years depending on species, trend strength, and study design.

In line with theoretical predictions (Rhodes and Jonzen 2011), I also found T_{min} was strongly correlated with the trend strength, variability in population size, and temporal autocorrelation (Fig. 3). Contrary to my prior expectations, I also found that T_{min} did not correlate with any life-history traits (Fig. 4). I initially hypothesized that species with longer lifespans or generation times may require a longer sampling period. This result could have been a result of at least two factors. First, the data I used may not include a diverse enough set of species with different life-history traits. Second, the question I poised, whether a population is increasing or decreasing, was specifically concerned with trends in population density over time. Therefore, life-history characteristics may be more important for other questions, like estimating species extinction risk (J. A. Hutchings et al. 2012). For example, Blanchard,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

Maxwell, and Jennings (2007) used detailed simulations of spatially-distributed fisheries to compare survey designs. They found that statistical power depended on survey design, temperature preferences, and the degree of population patchiness.

An important related question, is the optimal allocation of sampling effort in space versus time. In a theoretical investigation of this question, Rhodes and Jonzen (2011) found that the optimal allocation of sampling depended strongly on temporal and spatial autocorrelation. If spatial population dynamics were highly correlated, then it was better to sample more temporally, and vice versa. My work supports this idea as populations with strong temporal autocorrelation needed less years of sampling (Fig. 3). Morrison and Hik (2008) also studied the optimal allocation of sampling effort in space versus time, but used emprical data from a long-term survey of the collared pika (*Ochotona collaris*) found in the Yukon. They estimated long-term growth rates among three subpopulations over a 10-year period. They found that surveys less than 5 years may be misleading and that extrapolating from one population to another, even when nearby geographically, may be untenable.

Seavy and Reynolds (2007) asked whether statistical power was even a useful framework for assessing long-term population trends. They used 24 years of census data on Red-tailed Tropicbirds (*Phaethon rubricauda*) in Hawaii and showed that to detect a 50% decline over 10 years almost always resulting in high statistical power (above 0.8). Therefore, they cautioned against only using power analyses to design monitoring schemes and instead argued for metrics that would increase precision. For example, Seavy and Reynolds (2007) suggest improving randomization, reducing bias, and increases detection probability when designing and evaluating monitoring programs. I agree that power analyses should not be the only consideration when designing monitoring schemes. However, unlike Seavy and Reynolds (2007), my results indicate that longer than 10 years is often needed to achieve high statistical power.

This paper also has practical implications for the IUCN Redlist criteria. IUCN criteria A2 suggests that species that have experienced 30% declines over 10 years (or three generations) should be listed as vulnerable (IUCN 2012). However, for the populations I examined, this criteria may be too simplistic (Fig. 5). For many populations, the IUCN criteria suggest more years than necessary are required to assess a population as vulnerable (points below diagonal line in Fig. 5). Conversely, for other populations the IUCN criteria suggest sampling times that are less than the minimum time required for statistical power. This suggests that the IUCN criteria are probably too simplistic as the minimum time required does not correlate with generation time (Fig. 4).

The design of monitoring programs should include calculations of statistical power, the allocation of sampling in space versus time (Rhodes and Jonzen 2011), and metrics to increase precision. Ideally, a formal decision analysis to evaluate these different factors would be conducted to design or assess any monitoring program (Hauser, Pople, and Possingham 2006; McDonald-Madden et al. 2010). This type of formal decision analysis would also include information on the costs of monitoring. These costs include the actual costs of sampling (Brashares and Sam 2005) and the ecological costs on inaction (Thompson et al. 2000).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

4.1 Limitations

This paper has some limitations in determining the minimum time series length required. First, T_{min} is particular to the specific question of interest. An additional complication is that for the empirical approach, the subsampling of the full time series allows for estimates of power, but the individual subsamples are clearly not independent of one another. Further, estimates of T_{min} depend on chosen values of α and β (Fig. A4). In an ideal setting, a specific population model would be parameterized for each population of interest (McCain, Szewczyk, and Knight 2016). Then, model simulations could be used to estimate the minimum time series required to address each specific question of interest. Clearly, this is not always practical, especially if conducting analyses for a wide array of species as I do here. In addition, the statistical models suggest that T_{min} does not correlate with any life-history traits, at least for the question of linear regression (Fig. 4). Therefore, it is not possible to use these results to predict T_{min} for another population, even if the population is of a species with a similar life-history to one in the database used.

4.2 Conclusions

I used a database of 822 populations to determine the minimum time series length required to detect population trends. This goes beyond previous work that either focused on theoretical investigations or a limited number of species. I show that to identify long-term changes in abundance, on average 15.91 years of continuous monitoring are often required (Fig. 2). However, there is wide distribution of estimated minimum times. Therefore, it is probably not wise to use a simple threshold number of years in monitoring design.

In line with theoretical predictions (Gerrodette 1987), I also show that T_{min} is strongly correlated with the long-term population trend (i.e. rate of increase), variability in population size, and the temporal autocorrelation (Fig. 3). Contrary to my initial hypotheses, minimum time required did not correlate with generation time or any other life-history traits (Fig. 4). This result argues against overly simplified measures of minimum sampling time based on generation length (Fig. 5).

My work implies that for many populations, short time series are probably not reliable for detecting population trends. This result highlights the importance of long-term monitoring programs. From both a scientific and management perspective estimates of T_{min} are important. If a time series is too short, we lack the statistical power to reliably detect long-term population trends. In addition, a time series that is too long may be a poor use of already limited funds (Gerber, DeMaster, and Kareiva 1999). Further, more data is not always best in situations where management actions need to be taken (Martin et al. 2012; Martin et al. 2017). When a population trend is detected, it may be too late for management action. In these situations, the precautionary principle may be more appropriate (Thompson et al. 2000). Future work should examine other species, with a wider range of life-history characteristics. In addition, similar approaches can be used to determine the minimum time series length required to address additional questions of interest.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

5 Supporting Information

In the supporting material, I provide an expanded methods sections, additional figures, minimum time calculations for determining exponential growth, simulations with a more complicated population model, and the use of generalized additive models to identify population trends. All code and data can be found at https://github.com/erwhite1/time-series-project

6 Acknowledgements

ERW was partially supported by a National Science Foundation Graduate Fellowship. I would like to thank members of the Ecological Theory group at the University of California, Davis for their insight. I would also like to thank T. Dallas and E. Malcolm-White for their helpful comments.

7 References

Blanchard, Julia L, David L Maxwell, and Simon Jennings. 2007. "Power of monitoring surveys to detect abundance trends in depleted populations: the effects of density-dependent habitat use, patchiness, and climate change." *ICES Journal of Marine Science* 65 (1): 111–20.

Bolker, Benjamin M. 2008. *Ecological Models and Data in R*. 1st ed. Princeton, New Jersey: Princeton University Press.

Brashares, Justin S, and Moses K Sam. 2005. "How much is enough? Estimating the minimum sampling required for effective monitoring of African reserves." *Biodiversity and Conservation* 14: 2709–22. doi:10.1007/s10531-005-8404-z.

Cohen, Jacob. 1992. "A power primer." *Psychological Bulletin* 112 (1): 155–59. doi:10.1037/0033-2909.112.1.155.

Field, Scott A, Patrick J O Connor, Andrew J Tyre, and Hugh P Possingham. 2007. "Making monitoring meaningful." *Austral Ecology* 32: 485–91. doi:10.1111/j.1442-9993.2007.01715.x.

Gerber, L R, D P DeMaster, and P M Kareiva. 1999. "Gray whales and the value of monitering data in implementing the U.S. endangered species act." *Conservation Biology* 13 (5): 1215–9.

Gerrodette, Tim. 1987. "A power analysis for detecting trends." doi:10.2307/1939220.

Giron-Nava, Alfredo, Chase C James, Andrew F Johnson, David Dannecker, Bethany Kolody, Adrienne Lee, Maitreyi Nagarkar, et al. 2017. "Quantitative argument for long-term ecological monitoring." *Marine Ecology Progress Series* 572: 269–74.

Hatch, S A. 2003. "Statistical power for detecting trends with applications to seabirds monitoring." *Biological Conservation* 111: 317–29.

Hauser, Cindy E., Anthony R. Pople, and Hugh P. Possingham. 2006. "Should managed populations be monitored every year?" *Ecological Applications* 16 (2): 807–19.

Hughes, Brent B, Rodrigo Beas-luna, Allison K Barner, Kimberly Brewitt, Daniel R Brumbaugh, Elizabeth B. Cerny-Chipman, Sarah L. Close, et al. 2017. "Long-term studies contribute disproportionately to ecology and policy." *BioScience* 67 (3): 271–81. doi:10.1093/biosci/biw185.

Hutchings, Jeffrey A, Ransom A Myers, Verónica B García, Luis O Lucifora, and Anna Kuparinen. 2012. "Life-history correlates of extinction risk and recovery potential." *Ecological Applications* 22 (4): 1061–7.

IUCN. 2012. "IUCN Red List Categories and Criteria: Version 3.1." doi:10.9782-8317-0633-5.

Johnson, Paul CD, Sarah JE Barry, Heather M Ferguson, and Pie Müller. 2015. "Power analysis for generalized linear mixed models in ecology and evolution." *Methods in Ecology and Evolution* 6 (2): 133–42. doi:10.1111/2041-210X.12306.

Keith, David, H. Resit Akçakaya, Stuart H.M. Butchart, Ben Collen, Nicholas K. Dulvy, Elizabeth E. Holmes, Jeffrey A. Hutchings, et al. 2015. "Temporal correlations in population trends: Conservation implications from time-series analysis of diverse animal taxa." *Biological Conservation* 192. Elsevier B.V.: 247–57. doi:10.1016/j.biocon.2015.09.021.

Legg, Colin J, and Laszlo Nagy. 2006. "Why most conservation monitoring is, but need not be, a waste of time." *Journal of Environmental Management* 78: 194–99. doi:10.1016/j.jenvman.2005.04.016.

Magurran, Anne E, Stephen R Baillie, Stephen T Buckland, Jan Mcp Dick, David A Elston, E Marian Scott, Rognvald I Smith, Paul J Somerfield, and Allan D Watt. 2010. "Long-term datasets in biodiversity research and monitoring : assessing change in ecological communities through time." *Trends in Ecology and Evolution* 25: 574–82. doi:10.1016/j.tree.2010.06.016.

Mapstone, Bruce D. 1995. "Scalable decision rules for environmental impact studies : effect Size , type I , and type II errors." *Ecological Applications* 5 (2): 401–10.

Martin, Tara G., Abbey E. Camaclang, Hugh P. Possingham, Lynn A. Maguire, and Iadine Chadès. 2017. "Timing of Protection of Critical Habitat Matters." *Conservation Letters* 10 (3): 308–16. doi:10.1111/conl.12266.

Martin, Tara G., Simon Nally, Andrew A. Burbidge, Sophie Arnall, Stephen T. Garnett, Matt W. Hayward, Linda F. Lumsden, Peter Menkhorst, Eve Mcdonald-Madden, and Hugh P. Possingham. 2012. "Acting fast helps avoid extinction." *Conservation Letters* 5 (4): 274–80. doi:10.1111/j.1755-263X.2012.00239.x.

McCain, Christy Marie, Tim Szewczyk, and Kevin Bracy Knight. 2016. "Population variability complicates the accurate detection of climate change responses." *Global Change Biology* 22 (6): 2081–93. doi:10.1111/gcb.13211.

McDonald-Madden, Eve, Peter W J Baxter, Richard A. Fuller, Tara G. Martin, Edward T.

Game, Jensen Montambault, and Hugh P. Possingham. 2010. “Monitoring does not always count.” *Trends in Ecology and Evolution* 25 (10): 547–50. doi:10.1016/j.tree.2010.07.002.

Morrison, Shawm, and David S. Hik. 2008. “When? Where? And for how long? Census design considerations for an Alpine Lagomorph, the Collared pika.” In *Lagomorph Biology*, 103–13. Springer Berlin Heidelberg. doi:10.1007/978-3-540-72446-9.

Myhrvold, Nathan P., Elita Baldridge, Benjamin Chan, Dhileep Sivam, Daniel L. Freeman, and S.K. Morgan Ernest. 2015. “An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles.” *Ecology* 96 (11): 3109.

NERC Centre for Population Biology, Imperial College. 2010. “The Global Population Dynamics Database Version 2.”

Nichols, James D., and Bryon K. Williams. 2006. “Monitoring for conservation.” *Trends in Ecology and Evolution* 21 (12): 668–73. doi:10.1016/j.tree.2006.08.007.

R Core Team. 2016. “R: A language and environment for statistical computing.” Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Rhodes, Jonathan R., and Niclas Jonzen. 2011. “Monitoring temporal trends in spatially structured populations: how should sampling effort be allocated between space and time?” *Ecography* 34 (6): 1040–8. doi:10.1111/j.1600-0587.2011.06370.x.

Rueda-Cediel, Pamela, Kurt E Anderson, Tracey J Regan, Janet Franklin, and M Regan. 2015. “Combined influences of model choice, data quality, and data quantity when estimating population trends.” *PLoSONE* 10 (7): e0132255. doi:10.1371/journal.pone.0132255.

Seavy, Nathaniel E., and Michelle H. Reynolds. 2007. “Is statistical power to detect trends a good assessment of population monitoring?” *Biological Conservation* 140 (1-2): 187–91. doi:10.1016/j.biocon.2007.08.007.

Thompson, Paul M, Ben Wilson, Kate Grellier, and Philip S Hammond. 2000. “Combining power analysis and population viability analysis to compare traditional and precautionary approaches to conservation of coastal cetaceans.” *Conservation Biology* 14 (5): 1253–63.

Wagner, Tyler, Christopher S. Vandergoot, and Jeff Tyson. 2009. “Evaluating the power to detect temporal trends in fishery-independent surveys - A case study based on gill nets set in the Ohio waters of Lake Erie for walleyes.” *North American Journal of Fisheries Management* 29: 805–16. doi:10.1577/M08-197.1.

White, Easton R, John D Nagy, and Samuel H Gruber. 2014. “Modeling the population dynamics of lemon sharks.” *Biology Direct* 9 (23): 1–18.

Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev, and Graham M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R.* New York: Springer.

8 Supplementary material

Figure A1: (a) Population size of Bigeye tuna (*Thunnus obesus*) over time. The line is the best fit line from linear regression. (b) Statistical power for different subsets of the time series in panel a.

Figure A2: Output of generalized linear model with a Poisson error structure for predicting the minimum time required with explanatory variables of the absolute value of the slope coefficient (or trend strength), temporal autocorrelation, and variability in population size.

Figure A3: (a) Minimum time required to estimate change in abundance by biological class, (b) long-term trend (estimated slope coefficient) by class, (c) coefficient of variation in population size by class, and (d) temporal autocorrelation by class.

Table A1: Output of generalized linear model to examine time series characteristics as correlates of the minimum time required for determining long-term population trends.

Table A2: Output of generalized linear model to examine life-history trait correlates of the minimum time required for determine long-term population trends.

Table A3: Output of generalized linear model to examine both time series characteristics and life-history trait correlates of the minimum time required for determine long-term population trends.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017

Figure A4: Minimum time required to assess long-term trends in abundance for values of statistical significance (α) and power ($1 - \beta$).

Figure A5: The difference between minimum time estimates is the minimum time required to achieve 0.8 statistical power versus the minimum time required under IUCN criteria A2 to classify a species as vulnerable. Each point represents a single population, all of which saw declines of 30% or greater over a 10 year period. (a) Difference between minimum time estimates versus the coefficient of variation in population size. (b) Difference between minimum time estimates versus the generation length in years.

Figure A6: Distribution of the minimum time required in order to detect a significant trend (at the 0.05 level) in log(abundance) given power of 0.8.

Figure A7: Statistical power for different length of time series simulations for a lemon shark population in Bimini, Bahamas.

Figure A8: (a) Time series for Bigeye tuna (*Thunnus obesus*) with corresponding fitted GAM model in red and (b) statistical power as a function of the number of years sampled. The horizontal line at 0.8 indicates the minimum threshold for statistical power and the vertical line denotes the minimum time required to achieve 0.8 statistical power.

Figure A9: Distribution of the minimum time required in order to detect a significant trend (at the 0.05 level) in abundance according to a GAM model given statistical power of 0.8. The smoothing parameter was set to 3 for each population.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3168v3 | CC BY 4.0 Open Access | rec: 14 Dec 2017, publ: 15 Dec 2017