2017

Why most studied populations should decline

Methods: Portal data analysis

To test the site selection hypothesis, a time series with plot replicates was needed. However, in most ecological monitoring, there is a trade-off in sampling spatially or temporally. Typically, a long time series consists of a single population. We wanted to investigate the effect of biased sampling and therefore needed a long time series with multiple plot replicates.

We examined data from the Portal Project. The project consists of long-term monitoring of a Chihuahuan Desert ecosystem near Portal, Arizona, USA (Ernest et al. 2009). Since 1978, 24 individual replicate plots have been sampled. Monitoring includes ants, plants, and rodents. The experimental design, with replicates of plots, allows us to test ideas of sampling bias. Therefore, we explored biased subsets of the data and observed the effect of the sampling on assessing long-term trends.

We compared two subsets of data: 1) total abundance of the two initially (first 5 years) most abundant plots and 2) random selections of any pair of plots. This biased sampling allowed us to see the effect of only sampling the most common plots. We then used simple linear regression to estimate the population increase or decrease over time. We hypothesized that the initially most abundant plots should see significantly larger declines (i.e. more negative slope estimates) than the random subsets of plots.

We also examined the effect of removing the first five years of data. We hypothesized that by removing the first five years, that we should reduce bias in selecting only the most common sites.

Results: Portal data

We compared the predicted slope from linear regression for the two most common plots versus pairs of random plots. We found that the most common plots differ from pairs of random plots (Fig. 1). For 11 out of the 17 species, the slope from the most common plots were significantly different than choosing plots at random. For each case (except one) where the most common plots had a significantly different slope, the most common plots had a slope coefficient much more negative than random plot slopes.

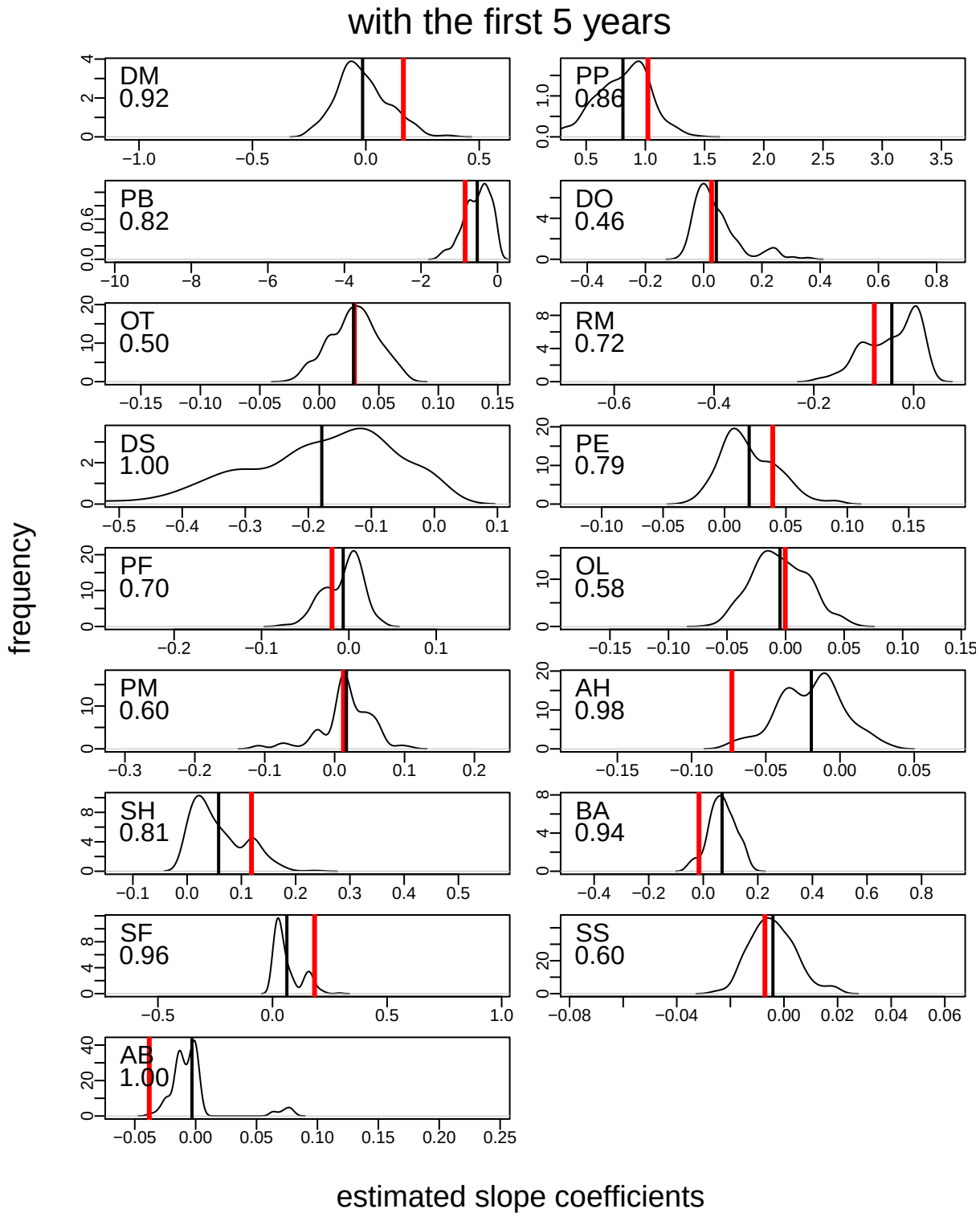

Figure 1: This figure denotes the subset of data where the first five years of data are included. Each plot represents a single species. Each plot has a distribution of slope values for random pairs of sites over time. THe vertical black line is the mean of this distirbution. The vertical, red line denotes the estimated slope for the two most abundant sites over time. The number in the top-right of the plot is the fraction of the distribution less than or greater than the slope estimate for the most common sites.

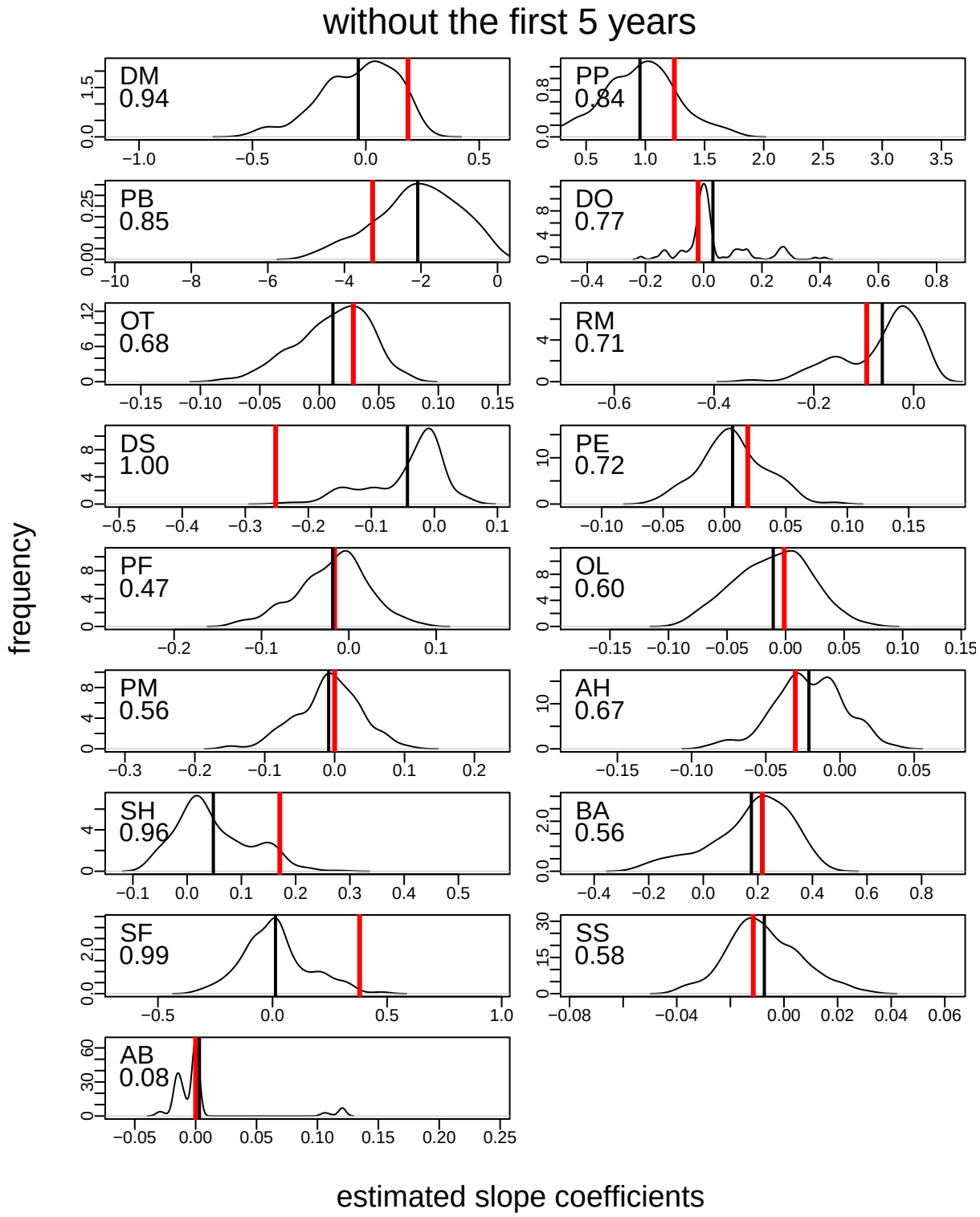

Figure 2: This figure denotes the subset of data where the first five years of data are not included. Each plot represents a single species. Each plot has a distribution of slope values for random pairs of sites over time. THe vertical black line is the mean of this distirbution. The vertical, red line denotes the estimated slope for the two most abundant sites over time. The number in the top-right of the plot is the fraction of the distribution less than or greater than the slope estimate for the most common sites.

We then removed the first five years of data. We hypothesized that this would help ameliorate some of the effect of choosing to sample only the most common plots. We found that fewer slopes from the common plots were statistically different from random slopes, although the effects still persisted (Fig. 2).

Alternatively, we can also just examine the species that saw declines over time. In Fig. 3 we examine the effect of removing the first five years of data for only these declining species. We see that removing the first five years of data either had no effect, or slightly reduced the difference between choosing random plots and choosing the most common plots (Fig. 3).

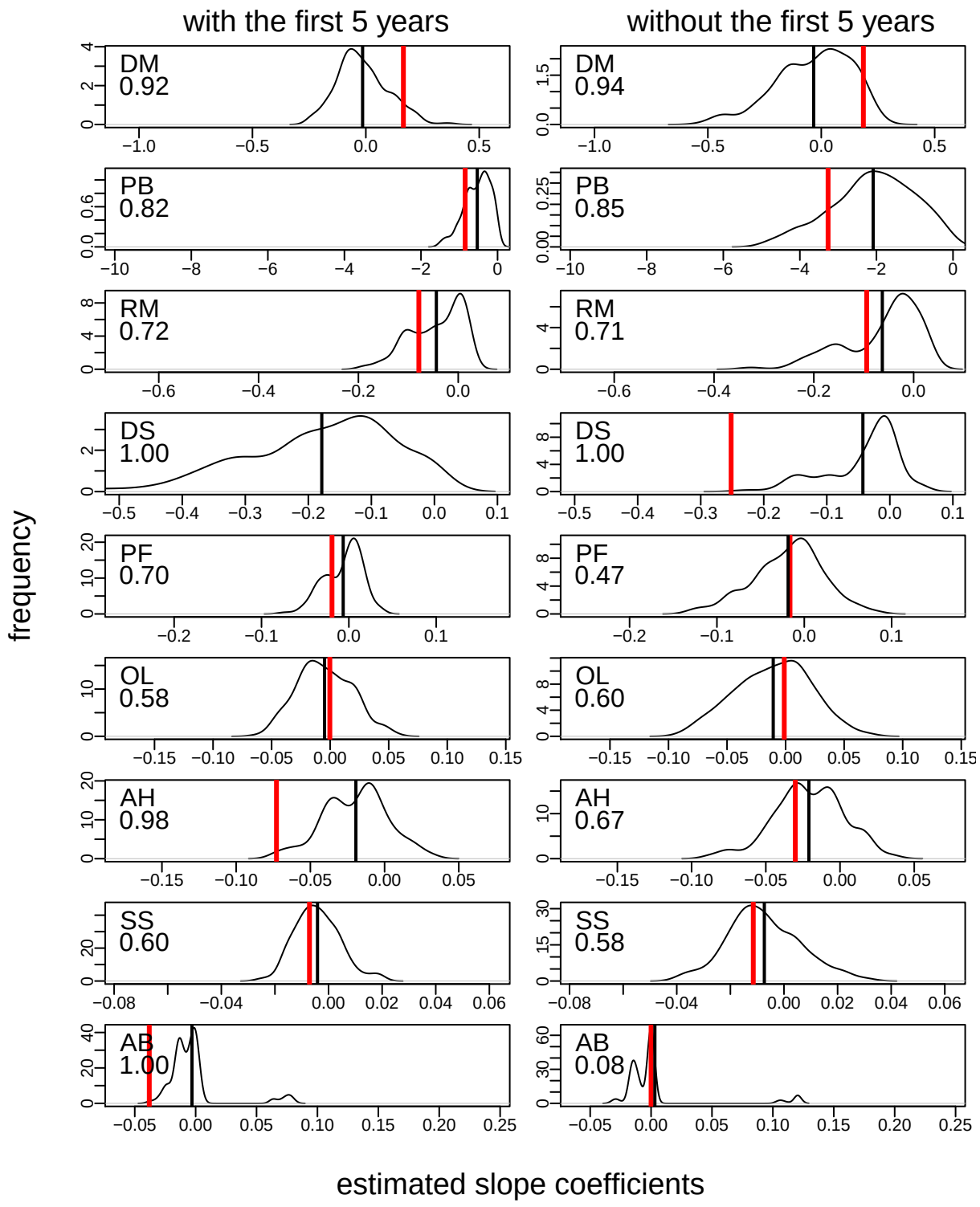

Figure 3: This figure denotes the subset of time series where the trend in population size was declining. Each row represents a single species. Each plot has a distribution of slope values for random pairs of sites over time. The vertical black line is the mean of this distirbution. The vertical, red line denotes the estimated slope for the two most abundant sites over time. The number in the top-right of the plot is the fraction of the distribution less than or greater than the slope estimate for the most common sites.

Methods: GPDD exploration

We also examined data from the Global Population Dynamics Database. This data does not have replicate plots like the Portal data. Instead, the database consists of time series for individual populations. We wanted to see the effect of sampling a population starting at a high point. The rational being that when initiating a survey you may start with a population at high abundance for logistical reasons.

We selected only time series with 40+ years of data to ensure high statistical power (White 2017). Then, for each population time series we examined two subsets of data: 1) sampling for 15 years starting at the time series high point, and 2) sampling for seven years on either side of the time series high point. The latter subset is to represent a situation where you remove some of the bias associated with only sampling populations at high abundance.

Results: GPDD exploration

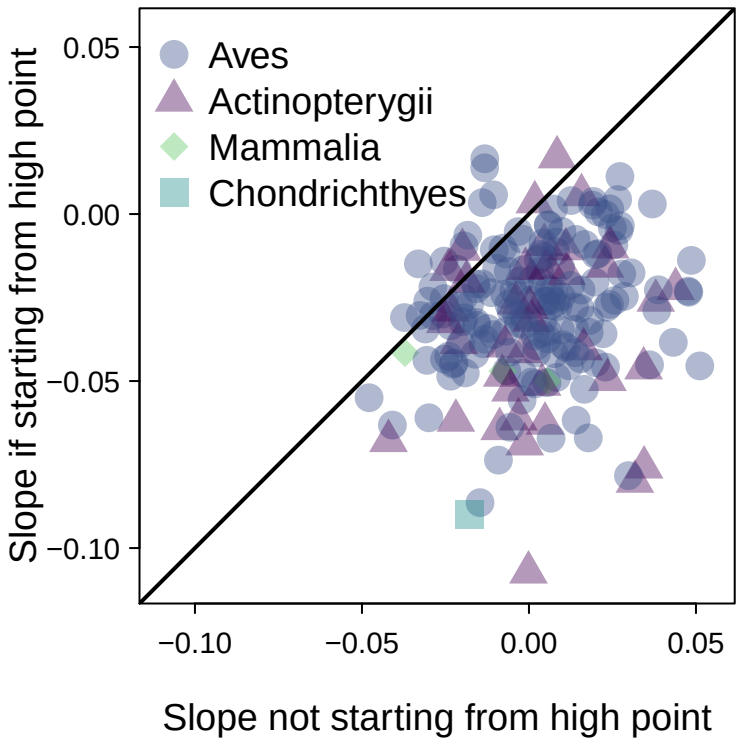

Figure 4: Slope estimate from linear regression for biased sample of starting at the high point in the time series versus not sampling at the high point. Any points below the identity line are situations where the slope estimate starting from the high point was less than that of sampling not starting from the high point.

We explored time series data for 202 populations of mammals, fish, and birds. For each time series, we examined two different subsets: either starting at the time series high point or not. As we hypothesized, the slope of a time series starting from a high point was typically more negative than situations not sampled starting from a high point (Fig. 4).

Potential explanatory variables (e.g. class, variance in population size, autocorrelation generation size) did not strongly predict which time series were more likely to be affected by biased sampling of high populations (Fig. 5).

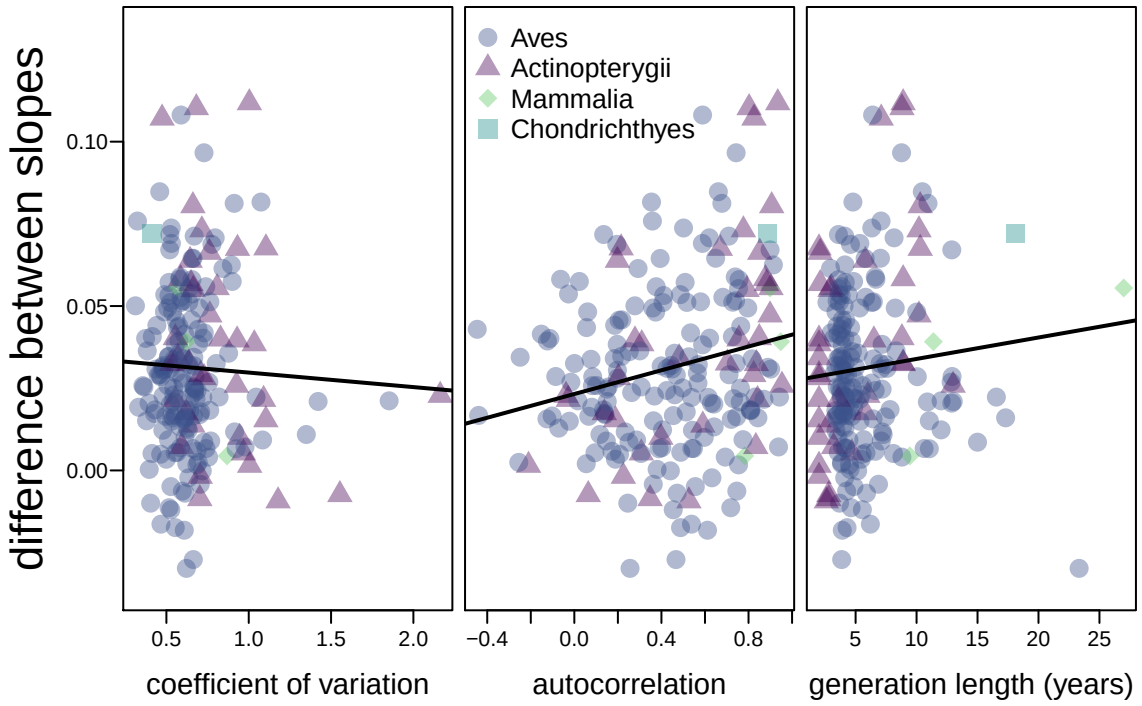

Figure 5: Slope estimate from linear regression for biased sample of starting to sample at the high point in the time series versus not sampling at the high point. Any points below the identity line are situations where the slope estimate starting from the high point was less than that of sampling not starting from the high point.

Removing initial years of time series

We also examined the effect of removing years from the beginning of the time series. The rationale here being that a census may have started in a year when a species was at a particularly high abundance. Therefore, the time series would start off artificially high. We subsampled each time series to remove initial years from the data. In other words, we examine the population size in years 1 through 15, then years 2 through 15, and so forth. We then examined how the estimated change in abundance (the slope coefficient) changed with more initial years removed. An example of this is shown in figure 6b. Here, a positive relationship between the trend estimate (slope coefficient) and number of initial years removed would indicate a situation where the initial years of the time series did in fact cause the abundance trends to be more negative.

We then examined the estimated relationship between trend estimate and the number of initial years removed for each declining species (Fig. 6c). Here, a positive value for slope would indicate that the initial years of a time series were artificially high and these caused larger estimated declines. We see that there is a large number of time series with slope values between 0.05 and 0.1 (Fig. 6c). The dark line is a null expectation of this data. To obtain this null distribution, we simulated time series where the initial years were not biased to be larger values.

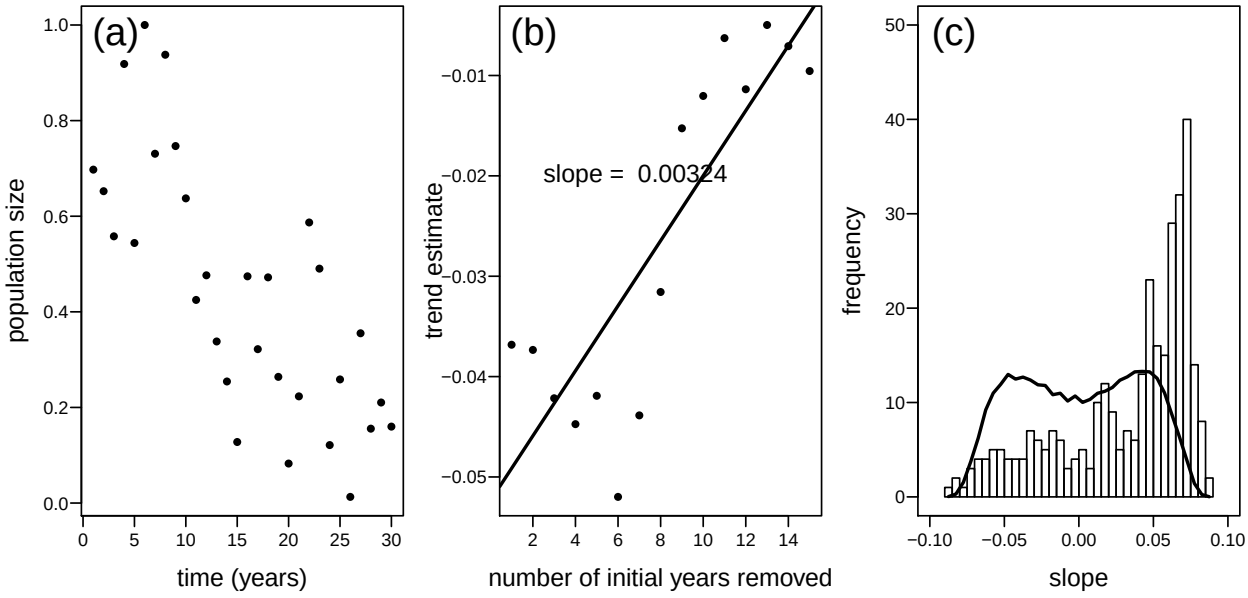

Figure 6: (a) Example time series from the GPDD database. (b) The trend in population size (i.e. the estimated slope coefficient) for different numbers of initial years removed from the data. In other words, we estimate the trend in abundance from years 1 through 15, then 2 through 16, and so forth. (c) Frequency of the slopes estimated to fit data in (b) for each species (n = 358). A positive value indicates the estimated trend in abundance over time becomes less negative when you remove more initial years. The dark line is the null distribution generated using simulated data.