Modeling Population Dynamics and Capturing Variability in Antarctic Penguins

A Dissertation presented

by

Emma J. Talis

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics & Statistics
(Computational Biology)

Stony Brook University

April 2023

Stony Brook University

The Graduate School

Emma J. Talis

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Dr. Heather Lynch – Dissertation Advisor
Professor, Dept. of Ecology & Evolution
IACS Endowed Chair in Ecology & Evolution, IACS
Affiliated Faculty, Dept. of Applied Mathematics & Statistics

Dr. Alan Tucker – Chairperson of Defense
Distinguished Teaching Professor, Dept. of Applied Mathematics & Statistics

Dr. Jie Yang
Associate Professor, Department of Family, Population and Preventive Medicine
Director, Renaissance School of Medicine Biostatistical Consulting Core
Affiliated Faculty, Dept. of Applied Mathematics & Statistics

Dr. Rafael D'Andrea – Outside Member
Assistant Professor, Dept. of Ecology & Evolution

This dissertation is accepted by the Graduate School

Celia Marshik

Interim Dean of the Graduate School

Abstract of the Dissertation

Modeling Population Dynamics and Capturing Variability in Antarctic Penguins

by

Emma J. Talis

Doctor of Philosophy

in

Applied Mathematics & Statistics
(Computational Biology)

Stony Brook University

2023

The Antarctic continent and surrounding Southern Ocean have been undergoing some of the fastest rates of warming on Earth. Penguins of the genus *Pygoscelis* are considered the bellwether sentinels of the Southern Ocean ecosystem on account of their sensitivity to changes in the environment. However, the extreme environmental conditions and logistical challenges of Antarctic research make data scarce on these three critical species, with penguin population time series being patchy at best, and nonexistent at worst. To understand how Antarctic penguin populations are changing, and thus, how climate change and other anthropogenic factors are affecting this globally important ecosystem, we must formulate accurate mathematical models for *Pygoscelis* penguin dynamics that accommodate their extreme stochasticity and develop methods for assessing their continent-wide trends.

In this dissertation, we first illustrate the difficulties in summing multiple log-normal distributions, which are often used to model animal abundance. This work was motivated by and illustrated using a recent study of global avian abundance; we find that the challenges are most acute when uncertainty is high, as is the case in many *Pygoscelis* time series. Next, we use the Adélie penguin (*Pygoscelis adeliae*) as a case study to show how empirical distributions for animal abundance may exhibit heavy-tailed dynamics. Using an age-structured model, we identify the causes of such heavy-tailed dynamics and link them to spatial patterns in Adélie food resource variability. We also quantify the impacts of demographic stochasticity, including skipped breeding, on simulated Adélie penguin time series to determine the mechanistic origins of the extreme variability observed in empirical *Pygoscelis* spp. time series. Additionally, we investigate potential null models for the Living Planet Index (LPI), a global biodiversity indicator, highlighting the effects of high interannual variability in population time series, like that observed for *Pygoscelis* penguins, on the index and associated null models. Lastly, we develop a *Pygoscelis* penguin-specific biodiversity

index, the 'Penguindex,' using the framework of the LPI, distilling 40-year population trends of *Pygoscelis* spp. penguins for the first time into a single, Antarctic-wide indicator for use by policymakers.

Contents

List of Figures

List of Tables

List of Publications

Peer-reviewed articles:

Talis, E. J., Che-Castaldo, C., Şen, B., Krumhardt, K., and Lynch, H. J. , "Variability, skipped breeding and heavy-tailed dynamics in an Antarctic seabird," *Journal of Animal Ecology*, 91.12, 2437-2450 (2022).

Talis, E. J., Che-Castaldo, C., and Lynch, H. J. , "Difficulties in summing log-normal distributions for abundance and potential solutions," *PLOS ONE*, 18.1 (2023).

Manuscripts under review:

Talis, E. J., Che-Castaldo, C., Hart, T. McRae, L. and Lynch, H. J. , "Penguindex: A biodiversity indicator for *Pygoscelis* species penguins identifies key eras of population change," *Polar Biology* (submitted February 2023).

Talis, E. J. and Lynch, H. J. , "The LPI is not inherently biased, but capturing stochasticity properly is key" *Nature Ecology and Evolution* (submitted February 2023).

List of Abbreviations

ACBR	**A**ntarctic **B**iogeographic **R**egions
AIC	**A**kaike **I**nformation **C**riterion
AP	**A**ntarctic **P**eninsula
BIC	**B**ayesian **I**nformation **C**riterion
CCAMLR	**C**onvention on the **C**onservation of **A**ntarctic **M**arine **L**iving **R**esources
CBD	**C**onservation on **B**iological **D**iversity
GAM	**G**eneralized **A**dditive **M**odel
JAGS	**J**ust **A**nother **G**ibbs **S**ampler
LPD	**L**iving **P**lanet **D**atabase
LPI	**L**iving **P**lanet **I**ndex
LPR	**L**iving **P**lanet **R**eport
MAPE	**M**ean **A**bsolute **P**ercentage **E**rror
MAPPPD	**M**apping **A**pplication for **P**enguin **P**opulations and **P**rojected **D**ynamics
MLE	**M**aximum **L**ikelihood **E**stimation
SSM	**S**tate **S**pace **M**odel
Spp.	**Sp**ecies (plural)

Acknowledgements

Chapter 1

Introduction

1.1 Antarctic penguins as sentinels of climate change

Increasingly, evidence has shown that ecosystems in the Southern Ocean are facing significant challenges with global implications [42, 186]. The Antarctic has undergone some of the fastest rates of warming on Earth [118, 156, 228, 227], with loss of sea ice and ocean acidification representing just two of the many associated consequences. The extreme environmental conditions of Antarctica beget a widespread scarcity of data throughout this critical region. Almost all of the Antarctic's terrestrial biodiversity exists in permanently ice-free areas that cover less than 1% of the continent's area [127]. However, most of these ice-free areas are difficult to access and, with very few long-term research stations on the continent [115, 202], sampling is generally restricted to the short summer months when travel to the Antarctic is most feasible. Thus data about Antarctic ecosystems, including animal abundance and distribution, are temporally, spatially, and taxonomically restricted. This paucity of data makes it difficult to track environmental changes in the Antarctic, including those related to climate change in addition to other anthropogenic influences including fishing, other resource exploitation, and invasive species, parasites, and diseases.

The genus *Pygoscelis* consists of three species of Antarctic penguins, Adélies (*P. adeliae*), Chinstraps (*P. antarctica*), and Gentoos (*P. papua*). *Pygoscelis* penguins are influential predators in the Southern Ocean ecosystem and are considered the bellwether

sentinels of the ocean on account of their sensitivity to changes in the environment [4, 25]. These penguins spend nine months of the year dispersed across the Antarctic, foraging on krill, crustaceans, or fish [45, 221]. As colonial breeders, *Pygoscelis* penguins congregate in densely-packed breeding colonies, consisting of up to hundreds of thousands of individuals [4], during the summer months [204]. *Pygoscelis* spp. penguins, especially Adélies, are highly site-faithful (or, 'philopatric'), returning to the same breeding colony (and even the same nest) year after year [6, 7]; though individuals may skip breeding in a given year in certain conditions [24, 98].

The colonial nature of their breeding during the austral summer makes *Pygoscelis* spp. penguins easier to monitor than most other species of Antarctic terrestrial or marine life. However, the logistical challenges of visiting the Antarctic still mar the ability of researchers to track these critical penguin species. Each species has hundreds of distinct breeding colony locations, spread across the sub-Antarctic islands and Antarctic Peninsula (all three spp.) and the entire coast of Antarctica (Adélie penguins). Though remote sensing methods have begun to show promise in monitoring *Pygoscelis* penguin abundance, most breeding abundance observations are ground counts, necessitating a team of researchers to visit each breeding site. Since visiting each breeding colony annually to hand-count penguin nests is infeasible, time series for *Pygoscelis* penguin abundance are patchy in both space and time. To understand how Antarctic penguin populations are changing, and thus, how climate change and other anthropogenic factors are affecting this globally important ecosystem, we must therefore formulate accurate mathematical models for *Pygoscelis* penguin dynamics. State-space models (SSMs) are one way of addressing some of these issues, allowing for the hindcasting, interpolation, and forecasting of animal abundance from these sparse time series [43, 114].

1.2 The classic modeling framework for animal population dynamics

Ecologists widely consider the population abundance of a single animal species to be log-normally distributed since populations grow multiplicatively [62, 87, 117]. MacArthur [146] first quantified this argument intuitively in 1960: assuming that the population growth rate of a species, with an abundance at time t of N_t and per individual growth rate of r_t, could be represented by the ordinary differential equation (ODE)

$$\frac{dN_t}{dt} = r_t N_t. \tag{1.1}$$

This ODE integrates to

$$\log(N_t) = \log(N_0) + \int_0^t r_t dt \tag{1.2}$$

(here and throughout we adopt the ecological notation, where $\log$ refers to the natural logarithm, $\ln$). MacArthur [146] argued that, since environmental factors might cause r_t to vary randomly over time, an application of the Central Limit Theorem would result in $\int_0^t r_t dt$, and thus $\log(N_t)$, being normally distributed. Dennis and Patil [62] recast this traditional multiplicative growth model as a stochastic differential equation to show that the log-normal distribution can be used as a model of a population with an environmental carrying capacity, not just a population undergoing unbounded exponential growth. Though there exist many other models for animal population dynamics, today the log-normal distribution is often used as the de facto model for population abundance [87]—a universal descriptor of abundance particularly when more detailed information is unavailable for a species.

SSMs are hierarchical models popularly used to model population dynamics, animal movement, and capture–recapture data [14, 43, 152]. SSMs are very flexible, allowing the natural variation in ecological processes to be modeled separately from any error associated with the observation process. The classic SSM for time series models of animal abundance involves these two main components: a *process model*, describing how the true abundance changes through time, and an *observation model*, describing the effect of sampling and any error associated with the measurement of true abundance [43]. By modeling the two sources of variability directly and individually, SSMs allow for the critical distinction between biological and demographic stochasticity and the error resulting from the sampling methodology, yielding more accurate analyses of ecological time series.

A simple process model for abundance in year t (N_t) would be

$$\log(N_t) = \log(N_{t-1}) + r + \epsilon, \qquad \epsilon \sim N(0, \sigma^2_{\text{proc}}), \tag{1.3}$$

with annual growth rate r, following directly from Equation (1.2), and ϵ representing the stochastic variation on the growth rate, frequently referred to as the "process error". This process model is often represented in the equivalent form

$$\log(N_t) \sim N(\log(N_{t-1}e^r), \sigma^2_{\text{proc}}). \tag{1.4}$$

Equivalently, this can be rewritten on the linear scale using a log-normal distribution:

$$N_t \sim \text{log-normal}(N_{t-1}e^r, \sigma^2_{\text{proc}}). \tag{1.5}$$

Layered on top of this model for the true abundance is usually a model for the observation or measurement process, wherein the observed abundance Y_t is related to the true abundance N_t. The simplest and most commonly used model for observation

error is

$$\log(Y_t) = \log(Y_{t-1}) + \omega, \qquad \omega \sim N(0, \tau^2_{\text{obs}}), \tag{1.6}$$

where ω represents the observation error and τ^2 represents the error variance. The use of the log-normal distribution to model the observation process often stems from observation errors being recorded as a percentage (e.g., $\pm 5\%$) of the counted abundance, a feature that can be closely approximated for reasonably small variances by the log-normal distribution.

Thus, models featuring either or both process and observation error lead to log-normally distributed abundance. The log-normal distribution has several advantages for modeling abundances, including its restriction to non-negative values and—unlike other strictly non-negative distributions such as the Poisson—the ability to tune the variance parameter to accommodate both under- and over-dispersion [131]. This log-normal framework has been used in SSMs describing the abundance (or, in other applications, the density or movement) of countless animal species [92, 152, 165, 174, 184], including *Pygoscelis* penguins in particular [39].

1.3 Capturing variability in Antarctic penguin time series

In addition to the data paucity brought on by the complicated logistics of Antarctic research, the population time series of *Pygoscelis* spp. penguins are also characterized by large interannual variability [39, 111, 240]. Ecologists have had challenges linking this population variability to environmental drivers, with existing studies only explaining a small percentage of these fluctuations to variables like sea ice concentration and sea surface temperature [40, 104]. Though the process error of an SSM can be adjusted

to accommodate increased interannual variability, such extreme unexplained stochasticity can make accurately modeling population time series difficult, hampering the ability to predict trends through time [23, 191].

In its simplest formulation, the modeling of population dynamics focuses on total abundance; however, ecologists often subdivide the population according to age, sex, or breeding status. In an age-structured model, a population is made up of individuals, distinguished by age, sexual maturity, or some other physical characteristic, and is assumed to evolve such that the structure of the population with respect to those individuals at a given time completely determines the population dynamics [1]. By explicitly representing biologically-meaningful transitions between states (e.g., survival and reproduction), such age-structured models can clarify which mechanistic and demographic drivers are responding to environmental conditions and influencing overall of population trends. Additionally, environmental information can be added to age-structured models to link trends to environmental variability. Since many colonially-breeding seabirds, including *Pygoscelis* penguins, are long-lived and display measurable population structure characterized by delayed sexual maturity, age-structured models provide a method to reveal relationships between demographic rates and population abundance and the associated variability [97, 107, 224].

The care needed to accurately analyze ecological time series for individual species, especially in the face of all the complexity involved, is often too demanding for policymakers who must make large-scale management decisions. In the Antarctic, the pipeline for making policy decisions is especially complicated, with a consensus-based governing framework controlled by the Antarctic Treaty Parties that incorporates guidance from the Scientific Committee on Antarctic Research (SCAR) and includes auxiliary policy-making bodies such as the Convention on the Conservation of Antarctic Marine Living Resources (CCAMLR). The (currently 56) nations that make up the Antarctic Treaty Parties regulate international relations, scientific cooperation, tourism,

conservation, and other Antarctic matters [196]. Now more than ever, it is critically important that these regulatory bodies can be presented with clear and concise descriptions of ecosystem trends, as CCAMLR continues the decades-long planning and implementation process for its Marine Protected Areas (MPAs) [22, 164] and the Antarctic Treaty Parties contend with both the increasing threats from climate change and the growing pressures of tourism.

To simplify the complexity of animal population dynamics, indices of ecosystem health—like the World Wildlife Fund's Living Planet Index (LPI) [8], the International Union for Conservation of Nature's Red List Index [230], and many others [33]—consolidate data from population time series into a single value or trendline to help policymakers understand how ecosystems are changing and what conservation efforts are needed. The LPI, a global biodiversity indicator for vertebrate species, has been adopted by the Conservation on Biological Diversity (CBD) [50, 51] as a metric to measure progress to the 2011 Aichi Biodiversity targets and the 2022 Kunming-Montreal Global Biodiversity Framework in addition to a myriad of other international policy documents [126]. However, the LPI, like most other indicators for global ecosystem change, has a data-scarcity problem that is perhaps most glaring for the Antarctic. The paucity of data available for *Pygoscelis* penguins in particular is exacerbated in the LPI, since the Antarctic research community has not made a strong effort to integrate these data into the databases underlying such biodiversity indicators. Therefore, there is currently no such easily-interpreted metric for ecosystem health and diversity available for use by Antarctic policymakers to make urgently important management decisions for this globally significant region.

1.4 Structure of the dissertation

The main focus of this dissertation is the development of more accurate methods for modeling and assessing the population dynamics of Antarctic penguins, with a particular focus on capturing the interannual variability in observed time series for abundance. The structure of the dissertation is as follows. Chapter 2 reviews the statistical issues involved in summing multiple log-normal distributions, often used to model animal abundance, and evaluates some alternative formulations that might be considered, particularly when abundance is estimated over multiple populations. Chapter 3 focuses on the possibility of heavy-tailed dynamics in *Pygoscelis* penguin population time series, focusing on Adélie penguins as a case study and identifying the best-fitting distribution for observed abundance before illustrating how a suite of candidate distributions can be modeled in a Bayesian state-space framework using JAGS ('Just Another Gibbs Sampler', [177]). Chapter 4 quantifies the impacts of demographic rates, including skipped breeding, on time series of Adélie penguin abundance simulated using an age-structured model, identifies the cause of heavy-tailed dynamics in simulated Adélie time series and, lastly, links these heavy-tailed dynamics to spatial patterns in Adélie food resource variability. Chapter 5 focuses on developing accurate null models for the LPI global biodiversity indicator, highlighting the effects of high interannual variability in population time series on the index and associated null models. Chapter 6 introduces a *Pygoscelis* penguin-specific biodiversity index, the 'Penguindex,' using the framework of the LPI, distilling 40-year population trends of *Pygoscelis* spp. penguins for the first time into a single global indicator for use by policymakers.

Chapter 2

Difficulties in summing log-normal distributions for abundance and potential solutions

2.1 Introduction

Log-normal distributions arise frequently in ecology and conservation, often as a means of modeling animal abundance and its attendant uncertainty, but their long right tails can make them difficult to use correctly. If the abundance of animal species is modeled by a log-normal distribution then the aggregate abundance across multiple populations (which could represent different populations of the same species or populations of different species) follows a distribution reflected by the sum of log-normal distributions. Unfortunately, unlike the normal or Poisson distributions, the log-normal distribution is not closed under addition; that is, the sum of log-normal distributions is not itself log-normally distributed. In fact, the sum of log-normal distributions does not follow a known probability distribution at all, and finding a closed form solution for the probability density function for the distribution of sums of log-normals is still an active area of mathematical research [19, 63, 67, 135]. Even more seriously, the median of the sum of log-normal distributions is not the sum of the medians of the constituent distributions. Because the median is the measure of central tendency

commonly used as a point estimate for skewed distributions, ecologists lack simple or well-known solutions for modeling animal abundance across multiple populations.

2.1.1 Use of the log-normal distribution in ecology and the modeling of animal abundance

Here we focus on the use of log-normal distributions in the modeling of animal abundance, using as a framework the classic description provided by Clark and Bjornstad [43] and subsequently developed by other authors [3, 60, 100]. The log-normal distribution has several advantages for modeling abundances, including its restriction to non-negative values and—unlike other strictly non-negative distributions such as the Poisson—the ability to tune the variance parameter separately from the mean, and thus to accommodate both under- and over-dispersion with respect to the mean. Thus the continuous log-normal distribution is often preferred to other distributions, including discrete distributions, to model the discrete variable of animal abundance. Two common scenarios leading to the use of log-normal distributions for abundance are models including normally-distributed random variation on growth rate ("process noise," which represents actual variability of the phenomenon of abundance) and models incorporating log-normally distributed observation error (representing the uncertainty in abundance estimates on account of the method in which estimates are determined). While these two sources of random variation are frequently included together in models of population dynamics [34, 43] [96, pp. 73-76 and 242-244], either one in isolation is sufficient to yield a distribution for abundance that is log-normal. It is important to note, however, that our investigation of this issue is agnostic to the method by which a statistical distribution modeling abundance has been derived.

There are many contexts in which ecologists might want to sum populations across multiple spatial or taxonomic units for the purposes of estimating a point estimate of abundance at larger scales. Most directly, a species may breed in spatially distinct

populations, as is common with seabirds, and an estimate of abundance or population trend may require summing these populations [39, 200]. Both metapopulation ecology [90, 91] and landscape demography [85] are concerned with the aggregated dynamics and persistence of spatially distinct populations. In conservation planning, it is a long-standing debate whether it is better to have a single large or several small reserves (the so called SLOSS debate, see [133, 173]), the comparison of which requires some way of summing populations across units. Moreover, ecologists are increasingly looking towards a portfolio approach to conservation, with due consideration for abundance and extinction risk across an ensemble of populations (e.g. [99]). Careful accounting of global abundance could also play an important role for assessment of a species for the International Union for Conservation of Nature Red List, for the designation of Important Bird Areas, or for protection under the US Endangered Species Act. It is important to note that here we focus on applications in which the total across two units should unambiguously equal the sum of the two units taken separately. Notably, the spatial scaling of species richness is not such a case, because species may be common to several units and the total richness (gamma diversity) is not simply the mathematical sum of the richness within each unit (alpha diversity). While analogous concerns may arise in this case as well, here we restrict our attention to the simplest scenario of summing quantities (such as the number of individual organisms) that are unique to each unit.

In its most general form, we will frame our analysis as follows: Consider a collection of different populations, each with abundance modeled by a log-normal distribution, representing the strictly positive number of animals ($X > 0$) within some spatial domain in which the animal is present. Here, we use the word 'population' to represent the smallest collection of individuals for which an abundance is estimated. While our analysis applies easily to other scenarios, in the context of our motivating case study [35] each population represents a different species and the aim is to estimate total abundance across all the species. For simplicity, we will refer to this as the

'global abundance'. We use the standard parameterization of the log-normal distribution $X \sim \text{Lognormal}(\mu, \sigma^2)$, where μ and σ represent the mean and standard deviation of the natural logarithm of the random variable (i.e, the abundance of one species), $\log_e(X)$. On the linear scale, the expected value (or mean) of abundance is given by $e^{\mu+\frac{\sigma^2}{2}}$, its median by e^{μ}, and its standard deviation by $\sqrt{(e^{\sigma^2}-1)e^{2\mu+\sigma^2}}$.

Estimates for abundance at larger scales should be consistent with the sum of estimates at smaller scales. In the absence of such scale-consistency, our inference becomes sensitive to the scale over which the assessment is made. This leads to a global abundance estimate that no longer corresponds to the individual abundances of each species, or estimates over a larger spatial scale that no longer corresponds to abundances within its constituent components.

2.1.2 Summing log-normal distributions

The global abundance across multiple species might be estimated in one of two ways, depending on whether the distributions are summarized or summed first. Here and throughout, we use 'summarize' to describe the process of obtaining a point estimate of abundance from a distribution while 'summing' abundance indicates we are mathematically adding together abundance estimates (in this case, across multiple species) to obtain an estimate of global abundance. As we will see below, the estimate of global abundance obtained by summing abundances may be an entire distribution, incorporating the attendant uncertainty in global abundance, or simply a point estimate.

First, one could summarize each species' abundance distribution to yield a species-specific point estimate to be added to the other point estimates. While unavoidable in contexts where only point estimates are available for each species, this approach precludes the propagation of uncertainty inherent to each individual species' estimates to the uncertainty appropriate to the aggregate. Alternatively, one could sum samples

from the individual abundance distributions and then summarize the resulting distribution to derive a point estimate for global abundance. In this way, the uncertainty of each individual species' abundance estimate propagates to the uncertainty in global abundance and a point estimate for global abundance can be easily obtained.

At first glance, this second method offers a deceptively straightforward solution to error propagation, involving nothing more than the addition of several distributions for the purposes of arriving at a distribution representing their sum. (As we illustrate in the next section, however, this approach is neither straightforward nor a solution.) While some would argue that the final distribution for abundance is sufficient and, in fact, superior to any measure of central tendency of that distribution, there are many contexts in which a point estimate for abundance is required. In such cases, there are several measures of central tendency that might be used for this purpose. Unlike the normal distribution, which is symmetric, the log-normal distribution is right-skewed, with a long right tail that pulls the mean to the right of the median. Because the mean, median, and mode are identical for the normal distribution, in practice the choice of the most appropriate statistic of central tendency is often irrelevant. For skewed distributions such as the log-normal, however, these three measures can be sharply divergent (Figure 2.1a).

While rarely discussed explicitly, implicit in the structure of many population models using the log-normal distribution is the assumption that e^{μ} represents the best estimate of true abundance [48, 187, 212]. There are several modeling techniques that allow for the distribution used for modeling abundance to be summarized by the point estimate e^{μ}. One method is to simply model abundance as a log-normally distributed variable and use the median (e^{μ}) as the point estimate (e.g.[172]). As it is often $\log_e(\text{abundance})$ that is the response variable being modeled, it is common to back-transform the mean of the normal distribution for $\log_e(\text{abundance})$, μ, to serve as a point estimate of abundance on the original (linear) scale as e^{μ} (e.g.[3]). Another approach, described by Hilborn and Mangel [96, pp. 73-76 and 242-244] and popular

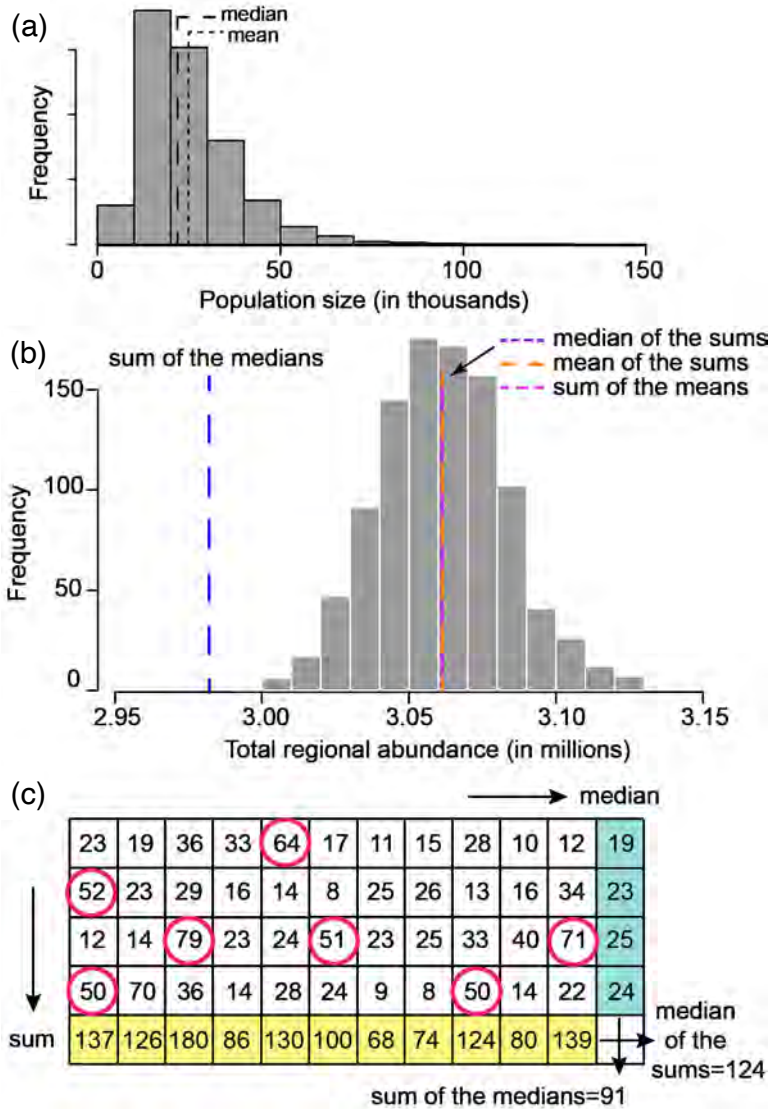

FIGURE 2.1: (a) The mean of the log-normal distribution is pulled right by the long right tail as compared to the median. (b) The distribution of sums of log-normally distributed abundance samples across 1,000 populations each distributed according to (a). The median of this distribution of sums is much larger than the sum of the medians of each individual population, since samples from log-normal distributions will include extreme values from the right-hand tails (circled pink in (c)). (c) This situation is illustrated with 4 populations and 11 samples from the distribution for each population.

in the fisheries literature [153, 162, 163, 211, 220, 233] involves taking the mean of the shifted log-normal distribution, Lognormal$(\mu - \sigma^2/2, \sigma^2)$, exploiting the fact that the expected value of this distribution is $e^{\mu-\sigma^2/2+\sigma^2/2} = e^{\mu}$. While these approaches yield equivalent point estimates for the true abundance for a single population, the situation becomes more complex when summing abundance distributions together.

2.1.3 The median of the sums is not the sum of the medians

As described above, a popular approach for summing multiple log-normal distributions is to simply sum samples from the individual distribution (i.e., each individual log-normal distribution modeling one species' abundance). An extension of the median as the measure of central tendency for each individual species uses the median of this derived quantity as the best estimate of global abundance. However, the entirety of the distribution for global abundance obtained by summing the samples from the individual species' distributions is larger than the sum of the medians of the constituent log-normal distributions (Figure 2.1b). Since the probability of the sum including an anomalously large value from the tail of one of the log-normal distributions (Figure 2.1c) grows as the number of species being summed increases, so does the difference between these two methods of estimating global abundance. As a result, summing abundances as a derived quantity for each draw from the distribution leads to an estimate larger than would have been obtained by summarizing the individual distributions for each population before summing. This distribution, derived from the sum across species, is not log-normally distributed (in fact, it has no well-described parametric form) and, being centered around the sum of the means of the constituent log-normal distributions, is uniformly too large to represent a reasonable global species sum.

Thus a simple sum of samples drawn from the constituent distributions yields a resulting distribution for global abundance that is too large, with no standard measures

of central tendency reflecting the sum of point estimates for the individual populations involved. For these reasons, we argue that the distribution for global abundance obtained by summing draws from the individual abundance distributions is as a whole unreasonable, having been pulled right by the long right tail of the constituent log-normal distributions. Nonetheless, practitioners summing across samples from the individual species-level distributions may not realize that the resulting quantity takes a distribution with unknown properties and moments that no longer reflect the true global abundance. We focus here on illustrating the difficulties inherent in summing log-normal distributions first using a pared down simulation study from which we can explore the conditions under which these challenges are most acute and, secondly, using a more concrete motivating example drawn from a recently published estimate of global avian abundance.

2.2 Simulation study: The extent of the problem

2.2.1 Summing independent and identical log-normal distributions

To demonstrate the complications inherent to summing log-normal distributions, we first considered a collection of $n = \{10, 100, 1000\}$ populations of animals. Each of these populations' abundance was modeled as $\log_e(N) \sim N(\mu, \sigma^2)$, where μ is the logged-median abundance and $\sigma = 0.20$ is the standard deviation on the $\log_e$ scale. Thus the n log-normal distributions each modeling an individual population's abundance are identical and independent with these parameters. We performed a suite of such simulations, with each simulation using a different value of μ for all n log-normal distributions. The range of μ values used, $2 \leq \mu \leq 12$, represents both small ($e^2 \approx 7$ individuals) and large ($e^{12} \approx 163,000$ individuals) populations. Setting $\sigma = 0.20$ yielded log-normal distributions each with an uncertainty of roughly -35% to $+50\%$ (taking the geometric mean, this is roughly $\pm 40\%$). Each individual abundance distribution

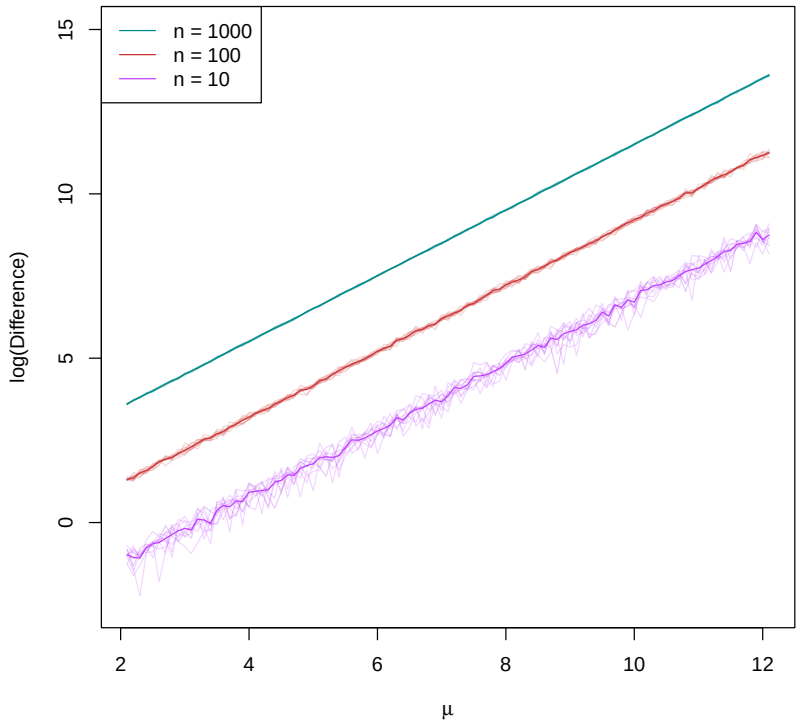

FIGURE 2.2: The logged difference in global abundance estimates (i.e., between the sum of the medians of the individual population abundances and the median of the distribution of summed of log-normal populations) for 10, 100, and 1,000 populations modeled using identical and independent log-normal distributions is plotted against the logged-median abundance μ. Solid lines represent the mean of the set of 10 iterations for each value of n.

consisted of $1,000$ random draws (i.e., realizations from the log-normal distribution). Following the reasoning above, we adopted the sum of medians of each population as the most direct reflection of global abundance and calculated the difference between this estimate of global abundance and that obtained by summing samples across populations to obtain a distribution of sums which is then summarized using the median. For each value of n, this difference was calculated 10 times for each value of μ and the mean difference was calculated over all 10 iterations. We found that the natural logarithm of this difference between the median of the distribution of sums for global abundance and the sum of the medians of individual populations grew linearly with

increasing logged-median abundance μ (Figure 2.2; see Appendix A.2 for mathematical derivation of linearity) and was larger for increasing numbers of populations. This difference grew sharply with increasing σ and was greater than 10% with highly uncertain abundances (not shown).

2.2.2 Summing correlated identical log-normal distributions

We next considered the case were individual populations were modeled using log-normal distributions that were identical but not independent. As above, we calculated the difference between the global abundance estimates. 100 related populations were modeled using identical multivariate log-normal distributions with $\mu = 6$ and $\sigma = 0.2$ and varying correlation coefficient $(0 - 1)$. This simulation was repeated for 100 iterations. As may be expected, the difference between the median of the distribution for global abundance (i.e., the distribution of sums) and the sum of the medians of individual populations shrunk as the correlation among populations increased (Figure 2.3). There was no difference between these two metrics when the populations were modeled by perfectly correlated log-normal distributions; in this extreme situation, the population collapsed to a single population with abundance modeled by a single log-normal distribution.

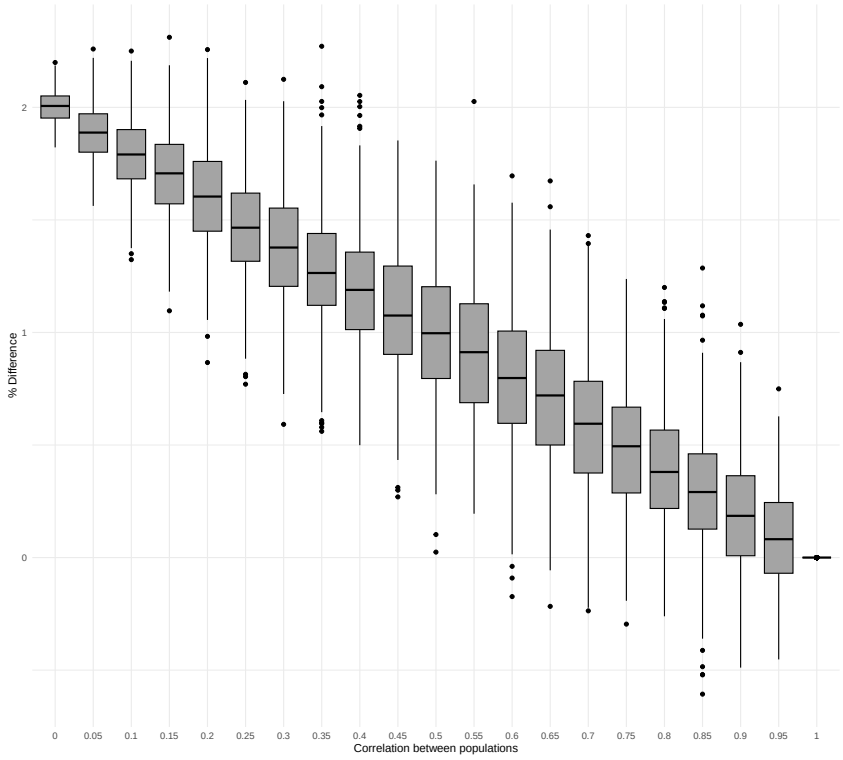

FIGURE 2.3: Box plots showing the percent difference in global abundance estimates from 100 simulations of 100 related populations modeled using multivariate identical log-normal distributions with varying correlation coefficient.

2.2.3 Summing independent and non-identical log-normal distributions

Lastly, we considered the case of summing non-identical log-normal distributions to understand the extent of the difference between the sum of the medians of the constituent distributions and the median of the distribution of sums when the individual log-normal distributions differ in mean and uncertainty. Specifically, we examined the case in which medium-sized populations known with reasonably high precision ($\mu_1 = 10$, median abundance $e^{\mu_1} \approx 22,000$, $\sigma_1 = 0.1$) are combined with large populations estimated with low precision ($\mu_2 = 12$, median abundance $e^{\mu_2} \approx 160,000$, $\sigma_2 = 0.4$). We calculated the difference between the median of the global abundance distribution and the sum of the medians of each population's abundance distribution

for different numbers ($n_1 = 1, ..., 25$ and $n_2 = 1, ..., 25$) of these different sized populations. We found that the percent difference in global population abundance estimates was more severe if the populations being summed were modeled using log-normal distributions that were not identical than compared to the case where the distributions were identical and independent (Figure 2.4). Further, the inclusion of more populations with high uncertainty caused the largest percent difference.

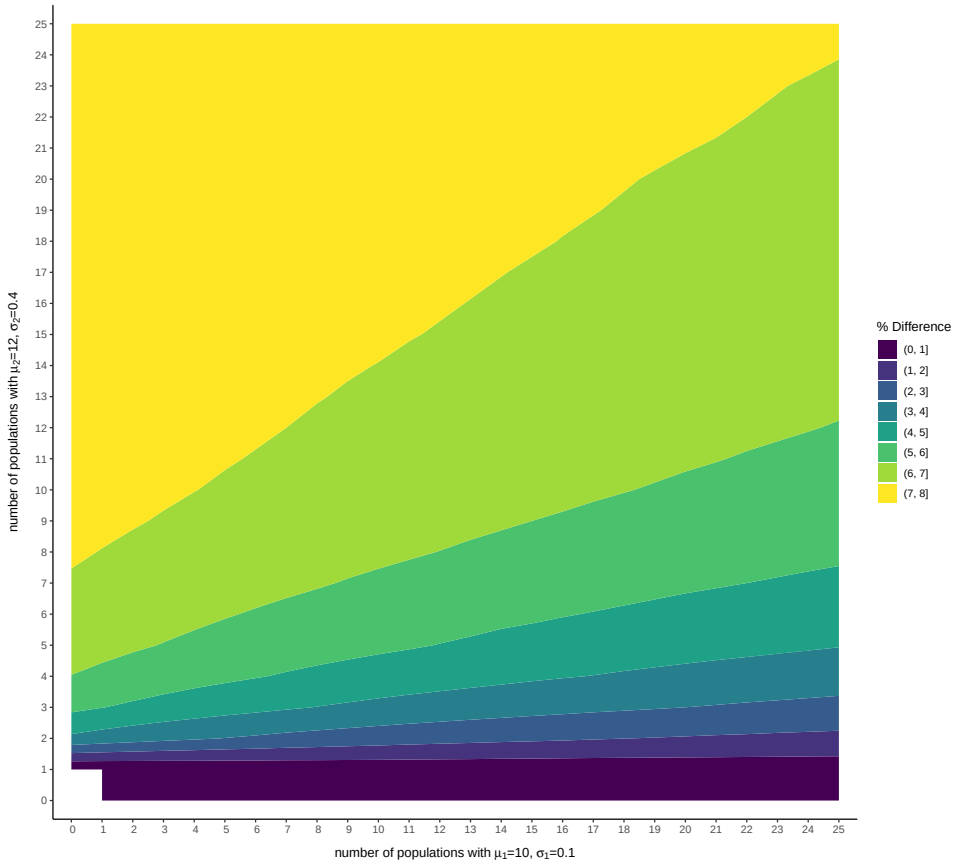

FIGURE 2.4: The average percent difference in global abundance estimates is shown for sums of n_1 identical and independent moderate-mean/low-variance log-normal distributions and n_2 populations modeled by identical and independent high-mean/high-variance log-normal distributions.

2.3 Global bird abundance study: A motivating example

We next used the dataset presented in the analysis of global bird abundance by Callaghan et al. [35] to illustrate these same ideas using a real dataset with significant conservation potential [147, 148, 158]. In the original study, the abundance of 9,700 individual bird species, each represented by a probability distribution, were summed to generate a distribution (and ultimately, a point estimate) representing global avian abundance [35]. In the context of this principal motivating example, we identified alternative models for species-level abundances yielding a distribution for global abundance closer to the value of the sum of the medians rather than the median of the sums. In this way, we obtained a scale-consistent estimate for global abundance, one that more closely aligned with the sum of the abundance estimates for each individual species (i.e., the sum of medians or, equivalently, $\sum_i e^{\mu_i}$) and captured the nature of the uncertainty for each species as well. It is worth noting that our use of the Callaghan et al. [35] data is not intended as a criticism of the original paper per se but simply offers an opportunity to illustrate a much more general issue in conservation biology using a dataset that is both recent and conservation relevant.

2.3.1 Reconstruction of Callaghan et al. data

In Callaghan et al. [35], the authors provide abundance estimates for each bird species summarized by the median and lower and upper bounds of the 95% CIs on $\log_{10}(\text{Abundance})$, which are assumed to be normally distributed (see Methods and Dataset S1 of Callaghan et al. [35]). We used the standard deviation for these normal distributions,

$$(\log_{10}(\text{lower limit}) + \log_{10}(\text{upper limit}))/2)/1.96 \tag{2.1}$$

and drew 10,000 random realizations from a normal distribution with mean equal to $\log_{10}(\text{median})$ and the calculated standard deviation. We then exponentiated (base 10) these realizations to yield a sample from the abundance distribution on the linear scale and fit a log-normal distribution (base e) to that sample, obtaining values for the parameters μ (the logged-median abundance, or mean on the $\log_e$ scale) and σ (the standard deviation of the $\log_e$ scale) of each log-normal distribution. We then obtained 10,000 random draws from each of these log-normal distributions, each used to model the abundance of a given bird species. This procedure allowed us to move from the original authors' [35] choice of a base-10 logarithm to a more traditional base-e logarithm and to generate a dataset that mirrors the original study's underlying population distributions [35].

Though our use of this dataset is intended for illustration purposes only, we wanted to confirm our reconstruction of the data used in Callaghan et al. [35], as doing so allows us to make direct comparisons between the estimates presented in the original study [35] and those resulting from the alternative summing and summarizing approaches addressed below. As such, we recreated each of the plots in Callaghan et al.'s [35] Fig. 2A and 2B (see Figures A.2 and A.3, respectively); we captured the data well, with a similar global species abundance distribution and individual distributions for species abundance. The average global population estimate over all 9,700 species was 5.2 million and the median estimate was 450,000, each equal to the estimate reported by Callaghan et al. [35].

We then generated a distribution for global abundance of birds by simply summing the distributions for individual abundance. To illustrate why choices regarding the distribution, summing approach, and measure of central tendency used all have consequences for point estimates of global abundance, we then calculated this sum using a suite of different approaches that might be used by ecologists looking for scale-consistent abundance estimates.

2.3.2 Alternative solutions

In their analysis, Callaghan et al. [35] circumvent the issue of the median of the sum of log-normal distributions being unequal to the sum of the constituent medians by sorting the samples within the distributions for each bird species before summing them. In other words, the smallest value in their distribution for global bird abundance represents the sum of the smallest values in the distribution for each bird species. As described by Callaghan et al. [35], this sorted summing "ensure[s] that the likelihood of values particular to each species correspond[s] with one another, therefore ensuring that the middle values correspond to those with the highest likelihood". Rephrased in the context of our analysis, this sorted summing ensures that the median of the resulting summed distribution is the sum of the medians. Using the log-normal distributions generated above, we followed this sorted summing procedure to obtain a second distribution for global abundance.

Another approach to ensure a scale-consistent distribution for global abundance is found in the aforementioned method described by Hilborn and Mangel [96, pp. 73-76 and 242-244], which shifts the log-normal abundance distribution for each species such that the mean—which has the desired additive property but is too large to use as the point estimate for abundance—replaces what was previously the median of the distribution. This model produces scale-consistent estimates for individual species' abundances (for which the point estimate is now the mean) and the total distribution for global bird abundance (the mean of which is equal to the sum of the means of each individual species' abundance). While this adjustment has been used previously [39, 181, 190] and is, in fact, quite standard within the fisheries literature [153, 162, 163, 211, 220, 233], there has been little discussion of its use in the context of population biology despite its utility in this context [39, 200]. We used this approach to obtain a third distribution for global abundance, using the parameters of the log-normal distributions calculated above with the Callaghan et al. data [35] to generate shifted log-normal

distributions as $N \sim$ log-normal $\left(\mu - \frac{\sigma^2}{2}, \sigma^2\right)$, and then summed them sample-wise. Note that, for this procedure and those that follow, we sum constituent distributions simply without sorting them first.

The constituent log-normal distributions for individual species abundance may also be replaced by the $(0, \infty)$ truncated normal distribution with the same mean and standard deviation (on the linear scale). Note that even with no change in μ or σ, the expectation of this zero-truncated distribution is greater than μ [18]. Accordingly, we obtained a fourth distribution for global abundance by summing zero-truncated normal distributions for each species as $N \sim \text{trunc}N\left(e^{\mu}, (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}, 0, \infty\right)$, where μ and σ are the parameters (on the $\log_e$ scale) of the uncorrected log-normal distribution calculated above and $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$ is the variance of that log-normal distribution on the linear scale.

Another correction involves replacing the log-normal distributions for abundance with a rectified normal distribution with mean e^{μ} and standard deviation σ. As opposed to the zero-truncated normal distribution, the rectified normal moves the statistical support for negative values to zero, leaving the median in most cases unchanged [203]. This rectification can take place at the level of individual populations (i.e., before summing) or can be done after summing. We present both methods below. These methods produce more scale-consistent distributions for global abundance by ensuring the median of the resulting distribution of sums is equal to the sum of the point estimates of the individual log-normal distributions representing single species' abundances. We generated our final two distributions for global abundance using normal distributions for each species as $N \sim N\left(e^{\mu}, (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}\right)$, where μ and σ are the parameters of the uncorrected log-normal distribution calculated above. For the rectify-then-sum procedure, we rectified the individual distributions by taking the maximum $\max(0, x)$ for each random draw x from the normal distribution and then summed the distributions sample-wise to obtain the distribution for global abundance. For the sum-then-rectify procedure, we sum the unrectified normal distributions sample-wise

and then took the maximum $\max(0, s)$ for each sum s in this global abundance distribution.

2.3.3 Performance of alternative solutions

As described above, in our re-analysis of the Callaghan et al. bird abundance data [35], we compared the summing of uncorrected log-normal distributions for individual species' abundance against five alternative approaches (Figure 2.5; note the $\log_{10}$ scale). Central tendencies used for the global abundance distributions are the following: median for sort-then-sum log-normals; mean for the shift-then-sum log-normals; mean for the truncate-then-sum normals; median for the rectify-then-sum normals; and median for the sum-then-rectify normals. Note that, in Figure 2.5, histogram bin widths underlying the kernel density plots shown are of equal width on the $\log_{10}$ scale used in this figure; as a result, the bulk of the area under the curve as displayed falls to the right of the distributions median. Plotted on a linear scale, the area under the curve would be equally partitioned to the right and left of the median. Here we have used the $\log_{10}$ scale both to match Figure 2.5b to Callaghan et al.'s original study [35] and because the differences between the methods are difficult to visualize on a linear scale.

The global abundance distribution obtained using the sort-then-sum log-normals method used by Callaghan et al. (shown in Figure 2.5b) is a direct reproduction of Fig. 2C in the original paper [35]. As was the motivation behind our study, the global abundance distribution obtained by simply summing uncorrected log-normal distributions for individual species' abundance (Figure 2.5a) was far from the sum of all species' abundance estimates in its entirety. Three alternative methods of generating a global abundance distribution had measures of central tendency close to the sum of the medians of the individual species' abundance distributions (cited by Callaghan et

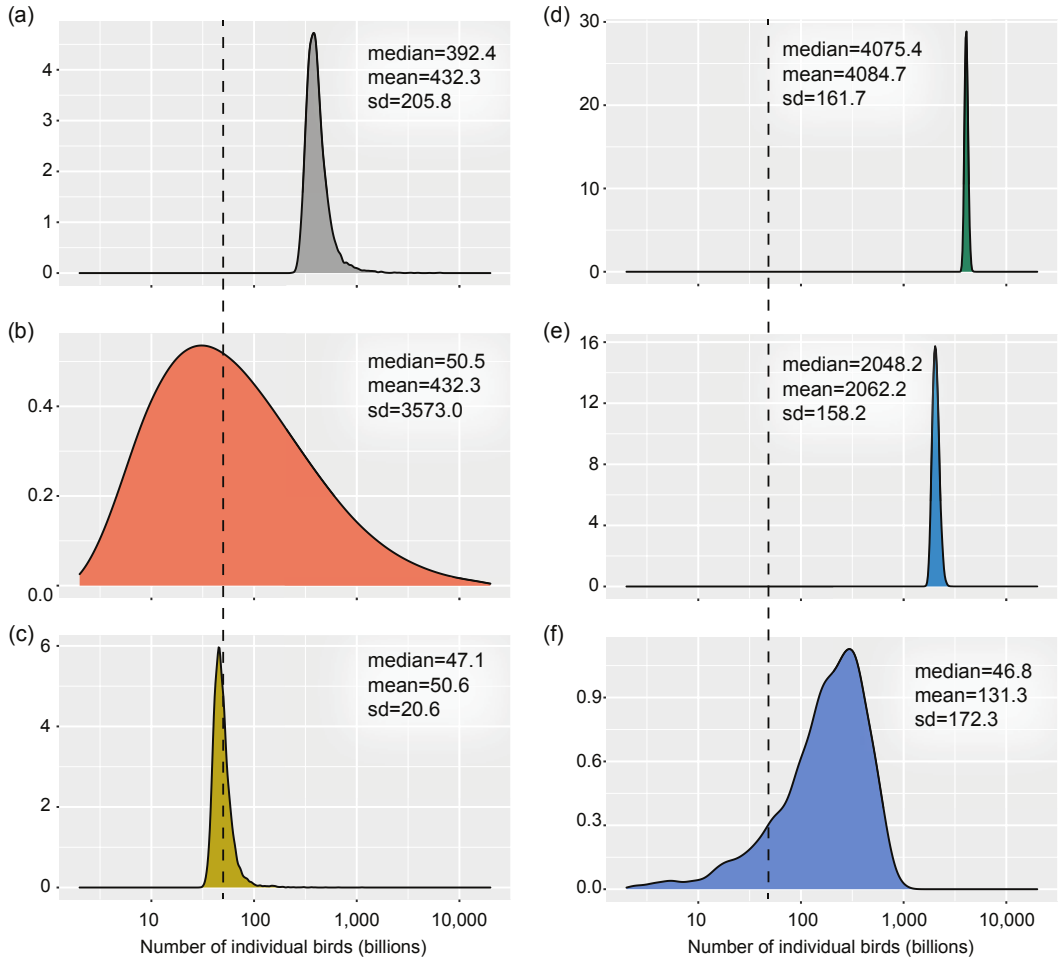

FIGURE 2.5: The total distribution of the number of individual birds in the world, calculated by summing all species-specific abundance distributions for 9,700 bird species using various corrections. The median, mean, and standard deviation (in billions) are shown for each distribution. The dotted vertical line indicates the sum of individual species abundance estimates (50 billion). (a) uncorrected sum of log-normals; (b) sort-then-sum of log-normals; (c) shift-then-sum of log-normals; (d) truncate-then-sum of normals; (e) rectify-then-sum of normals; (f) sum-then-rectify of normals. Details in the text.

al. as 50 billion [35]): sort-then-sum log-normal distributions (Figure 2.5b), shift-then-sum log-normal distributions (Figure 2.5c), and sum-then-rectify normal distributions (Figure 2.5f). Of these global abundance distributions [35], the one obtained using the sort-then-sum method (Figure 2.5b) had the largest standard deviation, 17 times larger than the standard deviation of the global abundance distribution obtained by

summing uncorrected log-normal distributions (Figure 2.5a) and more than 20 times larger than that of any other corrective method's resulting distribution. The global abundance distributions generated using truncate-then-sum (Figure 2.5d) and rectify-then-sum normal distributions (Figure 2.5e) were entirely larger than the sum of the medians of the individual populations, a problem even more extreme than for the global abundance distribution obtained by merely summing uncorrected log-normal distributions (Figure 2.5a).

Performance of alternative solutions in simulation study

We also calculated the percent difference between the sum of the abundance estimates for each individual species and the central tendency (either mean or median, see above) of the global abundance distribution generated by each of the alternative procedures presented above for the earlier simulation study. For each of the five procedures, 100 fictive populations are simulated using the given probability distribution with $\mu = \{2, 4, 6, 8\}$ and $\sigma = 0.2$ consisting of $10,000$ random draws. The distributions modeling each population's abundance are assumed to be identical and independent. The global abundance distribution was generated by summing these individual population abundance distributions, and an estimate for global abundance was obtained. This was repeated for a total of 100 ensembles for each procedure.

All five alternative approaches introduced above allowed for scale-consistent estimates of global abundance to be obtained in this case (Figure 2.6). The percent difference between the point estimate for global abundance and the sum of the medians of each individual population was negligible for each corrective procedure (compared to an average percent difference of 2% obtained summing the uncorrected sum of log-normal distributions). Across 100 ensembles, this percent difference was the least variable for the shift-then-sum log-normal procedure.

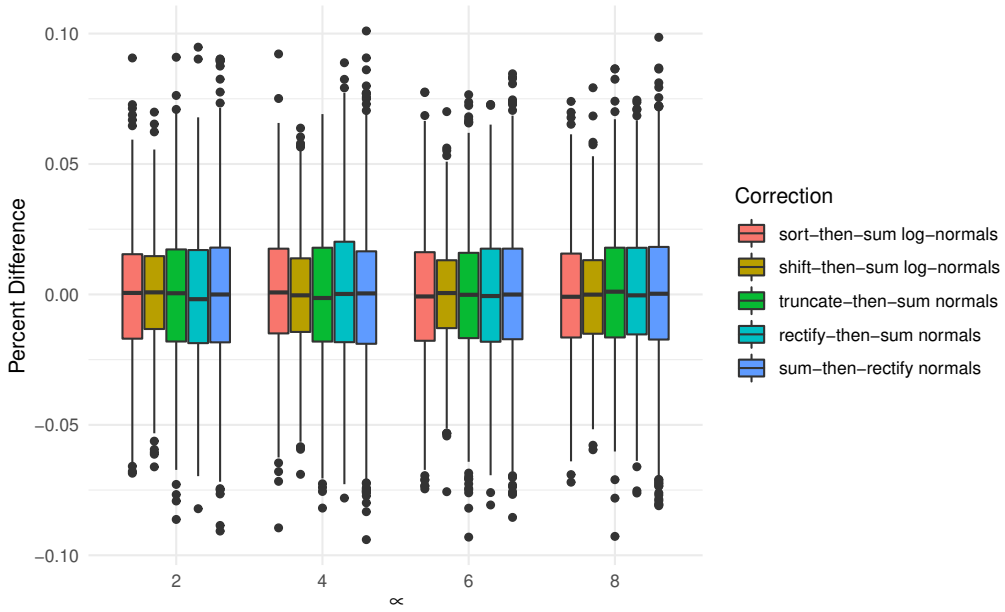

FIGURE 2.6: Box plots for the percent difference between the sum of empirical population medians (ne^{μ}) and the point estimate for global abundance obtained using each of the five corrective models for different values of logged-median abundance μ (box plot whiskers go to 1.5 times the interquartile range from the first and third quartiles).

2.4 Discussion

We have shown here the issues involved in summing log-normal distributions modeling abundance. These issues arise in a variety of contexts, regardless of the method by which a log-normal distribution modeling abundance has been derived. It is worth noting, however, that this distribution is often a posterior distribution estimated through a Bayesian analysis. While our examination of our findings includes some points specific to this Bayesian approach, it is important to remember that the issues of summing and summarizing distributions are generic and apply equally to both Bayesian and frequentist approaches.

2.4.1 The choice of point estimate

Our discussion of the challenges of summing log-normal distributions centers around the use of the median as the "best" measure of central tendency to summarize a log-normal distribution modeling abundance. It is worth reflecting on the advice provided to ecologists for this choice. The mode is, by definition, the value of the distribution that is most likely. In a Bayesian analysis, the mode is also the value most similar to the maximum likelihood estimate when the prior is uninformative. The mean and median minimize the sum of squared error and the sum of absolute error, respectively. These basic facts, however, are unhelpful for most practitioners in choosing an appropriate point estimate.

Perhaps surprisingly, ecological modeling textbooks provide scant advice on the matter and are largely agnostic on the choice between mean, median, and mode when deciding how to summarize a distribution, presenting them as largely interchangeable choices [80, 140]. Gelman et al. [80] present one example in which the distribution of a parameter includes unphysically negative values (akin to a normal distribution for abundance) and note that "the marginal posterior mean is not always a good summary of inference about a parameter…because the posterior mean includes the cases where [the parameter] is negative." Lunn et al. [140] touch on this briefly as well, noting that "[t]here have been limited 'guidelines' for reporting Bayesian analyses" and that "[w]here possible, full posterior distributions should be given…particularly for skewed distributions." While we agree that the presentation of the full distribution is important, it is no substitute for providing a concrete point estimate of abundance in conservation contexts, particularly when providing information to policymakers without specialized statistical knowledge.

Finally, as noted above, the uncertainty reflected in a distribution for abundance

often arises from observation error. Using the median as the point estimate for abundance implies that under-counts and over-counts are equally likely. Such an assumption might be reasonable for many sampling methodologies. Additionally, the mean ($e^{\mu+\sigma^2/2}$) changes as a function of a survey's precision. Thus, when using the mean as the point estimate of true abundance, a stable population surveyed over time would appear to decline as an artifact of declining measurement error over time (as might happen, for example, with improved survey methods). The use of the median avoids this logical inconsistency whereby different survey processes would change the underlying population abundance. For both of these reasons, the use of the median is often preferred over the mean when observation error is involved.

2.4.2 Callaghan et al.'s approach

The propagation of uncertainty from several log-normally distributed populations to a distribution representing their sum is a surprisingly difficult task that has garnered far too little attention given the consequences for conservation biology [85, 90, 122, 173]. The approach taken by Callaghan et al. [35] to sort the samples making up each individual species' log-normal abundance distribution before summing is, to our knowledge, an unusual choice. While this summing procedure did generate scale-consistent estimates for global abundance, it did so at the expense of the resulting distribution's variance. The standard deviation of the global abundance distribution became grossly exaggerated under the sort-then-sum procedure, with a median of 50 billion, a mean of 431 billion, and a range of 256 trillion (Figure 2.5b). This degree of uncertainty will render the estimates for global abundance useless in many conservation contexts, as was noted by critics [183].

Earlier concern regarding Callaghan et al.'s [35] methodologies, however, focused on the acutely uncertain and often biased constituent distributions modeling individual species' abundance [36, 183]. Our analysis identifies an important and previously

unrecognized driver of uncertainty in their final distribution for global bird abundance (the sort-then-sum procedure), and we have found that alternative approaches result in distributions for global bird abundance with dramatically lower uncertainty. The authors' choice of the sort-then-sum procedure and the associated uncertainty is not discussed in detail in the original study, nor is the difference between the mean and median as alternative point estimates. While modifying the method by which abundance distributions are summed will not address all of the concerns raised, it is important to separate out sources of uncertainty so that each can be considered thoughtfully and minimized where possible.

2.4.3 The choice of alternatives: Context matters

In our re-analysis of the global bird abundance data [35], two alternative models (i.e., replacing the log-normal distributions modeling individual species abundance with shifted log-normal distributions or with normal distributions that are rectified after summing) generated distributions for global abundance that were consistent with the sum of the estimates from the individual distributions for species abundance. Though they produced similar point estimates (using the mean and median, respectively), these two methods generated different distributions for global abundance (Figure 2.5c,f). The distribution obtained using the shift-then-sum log-normal distributions procedure had the smallest variance of any global abundance distribution, with a standard deviation about 10 times smaller than that of any other. The distribution obtained using the sum-then-rectify normal distributions procedure had a median close to the sum of the individual species' abundance estimates but the mean of the distribution is pushed right after the distribution is rectified following summing. Additionally, the global abundance distribution generated by summing zero-truncated normal distributions for individual species' abundance (Figure 2.5d) had a mean (and median) much larger than even the distribution obtained using raw log-normal distributions

on account of the increased expected values of the constituent truncated distributions [18]. Likewise, the global abundance distribution obtained using the rectify-then-sum normal distributions procedure (Figure 2.5e) was entirely too large, since the individual distributions representing species' abundances were heavily skewed by the rectification process. Clearly, these two procedures led to extreme scale-inconsistency for the Callaghan et al. data [35]. However, these issues may be less severe in a context where fewer, less abundant, or less uncertain individual species are being summed. In the case that species with more modest abundance uncertainties are modeled using identical and independent distributions, our simulation study showed that all alternative methods presented here produced scale-consistent point estimates for global abundance (Figure 2.6). Thus, while using shifted log-normal distributions worked well in both contexts shown here, the suitability of these alternative procedures should be assessed carefully for specific applications.

While we focus here on the use of log-normal distributions in the modeling of animal abundance, which is an easily conceptualized problem with clear conservation importance [3, 35, 48, 90, 99, 133, 173, 197, 207], there are many contexts in which ecologists might be trying to sum unique log-transformed items or measures across multiple temporal, spatial, or taxonomic scales. In addition to animal abundance, examples include quantities like biomass, mortality, and rainfall [17, 71, 113]. The use of the sort-then-sum log-normals, shift-then-sum log-normals, truncate-then-sum normals, rectify-then-sum normals, and sum-then-rectify normals procedures provided the scale-consistency that is desired in some scenarios. It is of the utmost importance, however, to note that each of these approaches modified the distribution of the sum in ways that can impact inference if the populations are very small, the uncertainties are very large, or if the analysis requires a careful consideration of distribution statistics other than the mean or median. The choice among these approaches depends on the specific application and whether inference hinges only on a measure of central tendency or whether other statistics, such as specific quantiles, may be needed. While

the use of shifted log-normal distributions works well in both extreme examples presented here and has been used previously [39, 181, 190, 200], it should not be adopted without thoughtful consideration of its impact in the context of the specific conservation question being asked.

2.4.4 Data-deficient species

As we have shown, the use of the log-normal distribution to model abundances caused the most significant challenges for summing when population estimates were highly uncertain, exactly the one might face with highly data-deficient species. Even for species that have been surveyed in detail, the uncertainty surrounding abundances can become very large over an extended period of time without observation data, since, as described by Clark et al. [43], “as the interval between observations widens, the variability contributed by process error increases correspondingly". In these cases, the most current understanding of abundance for any given population may be highly uncertain, and even small numbers of populations with such imprecise abundance estimates can contaminate global abundances unless care is taken to carefully choose the summing procedure and the measure of central tendency used as the reported point estimate. We also showed the challenge of summing abundances over multiple populations was reduced, but not eliminated, when abundances were correlated across populations. While such correlations may be present to a modest extent in a spatial context, in which neighboring populations may have correlated errors, such correlations would be absent in a context like that presented by Callaghan et al. [35]. In any case, it was only in the most extreme and highly unrealistic situation of populations with perfectly correlated errors that this difference actually collapsed to zero; in such a scenario, the populations are functionally one single population.

2.5 Conclusions

At its heart, our analysis calls into question under what conditions the log-normal distribution is appropriate for modeling animal abundance, particularly in cases where the distribution represents a Bayesian posterior distribution directly interpreted as the degree of belief one has in different values of abundance as true abundance (these more philosophical issues are discussed in Appendix A.3). Our choices of distributions for modeling animal abundance are extremely limited; our wish list for the ideal distribution includes one that is discrete, non-negative, closed under addition, and has a variance that can be tuned separately from the mean. We are aware of no parametric distribution that satisfies all four of these criteria [178], and thus it is incumbent on ecologists to select distributions whose shortcomings have the least impact on the problem at hand. In fact, many popular distributions are both skewed and not closed under addition, including the negative binomial distribution, and these also require considerable care in their use (see Appendix A.4). While the log-normal distribution emerges naturally from the exponential growth process and its non-negativity is convenient, its right skew can be severe when uncertainty is large and its use for conservation relevant analyses should be undertaken with care and careful consideration. The accurate estimation of abundance is important in its own right and we hope that our analysis will encourage ecologists to think carefully about how their choices in modeling abundance across different spatial or taxonomic scales may impact the final estimates provided to stakeholders.

Chapter 3

Heavy-tailed distributions in animal population modeling

3.1 Introduction

Time series modeling often involves two components, a process model (Equation 1.3) that describes the true underlying dynamics linking the abundance in year $t-1$ to the abundance in year t, and an observation model (Equation 1.6) that captures the observation process and any errors associated with the measurement of true abundance. We concern ourselves here with the process model. A simple process model for N_t (the abundance in year t) might follow Equation (1.3), which can be equivalently framed as

$$\log\left(\frac{N_{t+1}}{N_t}\right) = r + \epsilon, \qquad \epsilon \sim N(0, \sigma_{\text{proc}}) \tag{3.1}$$

where r represents the annual growth rate and σ_{proc} represents the process error (the variation in true abundance driven by unmodeled processes) [43]. This represents an exponential growth model, with the mean of the abundance in year t (N_t) depending on the previous year's abundance (N_{t-1}). Usually, the process model assumes that logged abundance follows a normal distribution (as in Equation 3.1); however, recent work [11] suggests that the normal distribution may not accurately fit the tails of some

empirical distributions for animal population growth rates, and thus has the potential for underestimating the probability of extreme population fluctuations.

Prior to the 2008 financial crisis, Nassim Taleb published *The Black Swan* [214], in which he warned of black-swan events: unpredictable and statistically improbable events that can nevertheless occur, often with extreme consequences. These black swans, though rare, have a disproportionately large role in the dynamics of the systems in which they occur. For example, large, unpredictable swings in the stock market account for the majority of financial earnings and losses [206, 214]. Additionally, the loss of life in the largest wars in history far outweighs that of all other wars [182]. In epidemiology, superspreading events occur when an infected individual infects an unusually high number of others [76, 134]; these events have been shown to drive outbreaks of several infectious diseases including HIV [154] and respiratory illnesses like SARS [199] and COVID-19 [10, 130]. Animal mass mortality events can also kill off millions, or even billions, of individuals at once [68].

Though Taleb's original theory espoused these black-swan events as unpredictable, many investigations—particularly in quantitative finance—have focused efforts on developing models that better account for these extreme events. One theory views these black swans as tail events. Instead of these extreme events being viewed as exceptional in some way, they are viewed simply as more extreme versions of standard events, belonging to the same statistical distribution. In this way, developing a more accurate model for the time series is a matter of finding an accurate distribution for population abundance, particularly one with tails heavy enough to accommodate these black swans [2, 11, 103, 134].

Because of their disproportional effect on the systems in which they occur, these black-swan events, sometimes referred to in ecology as 'catastrophes', are one of the most important considerations in predicting the persistence and viability of animal populations [29, 149]. Anderson et al. [11] studied extreme population-level events in over 600 animal species, finding evidence for the presence of black-swan events in

about 4% of the animal populations studied, and most frequently in birds.

Understanding process error is both theoretically important and critical for the imputation of missing data and forecasting of population viability under climate change. A particular example is that of the Adélie penguin (*Pygoscelis adeliae*), a colonial seabird distributed across the Antarctic coast. Changes in abundance of Adélie penguins have informed researchers' understanding of how climate change is affecting the Southern Ocean ecosystem [4, 39]. However, due to the remote nature of the Adélie habitat, it is often difficult to consistently census all known Adélie colony sites [141]. Thus it is important to be able to accurately model a time series of Adélie penguin abundance in order to predict continent-wide trends if the species is to be used as a climate change indicator. On account of the growing evidence for black-swan events in many animal populations, it is important to employ an accurate distribution of process change in any time series analysis of animal abundance.

Here we consider many alternative distributions to the normal distribution in Equation (3.1) to allow for a higher probability of extreme events. We first describe the distribution of empirical Adélie penguin population growth and fit close several alternative distributions to this data to identify the most promising candidates for an Adélie penguin process model. Next we illustrate how several of these candidate distributions can be modeled in a Bayesian framework for future modeling work for any animal species. In particular, we highlight the differences between employing these more exotic distributions by using built-in JAGS functions—which are limited in selection and can be parameterized quite non-intuitively—and manually defining likelihoods using either the 'zeros' or 'ones' tricks.

3.2 Heavy-tailed distributions for Adélie penguin abundance

To identify promising distributions to model Adélie penguin abundance, we first characterize the shape of the distribution of empirical Adélie growth to determine if the standard log-normal process model is in fact missing density in the tails (and thus underestimating the probability of extreme events). We use data from all 271 Adélie penguin breeding populations between 1970 and 2019, collected by the Mapping Application for Penguin Populations and Projected Dynamics (MAPPPD; [102]) and calculate the the logged change in abundance between consecutive years for which both years had measured counts ($\log(\frac{N_{t+1}}{N_t})$, see Equation 3.1). We consider alternatives to both the normal distribution in Equation (1.4) and the log-normal distribution in Equation (1.5), fitting these alternative distributions to $\log(\frac{N_{t+1}}{N_t})$ and $\frac{N_{t+1}}{N_t}$, respectively.

3.2.1 Quantifying heavy-tailedness

To determine if a normal distribution is the appropriate distribution for logged Adélie penguin abundance (and thus a log-normal distribution for abundance on the linear scale), we first calculate the skewness and kurtosis ('tailedness') of the empirical distribution for $\log(\frac{N_{t+1}}{N_t})$ to understand the shape of this data and find candidate distributions for fitting. A skewness-kurtosis plot (e.g., the one proposed by Cullen and Frey [56]) plots skewness and kurtosis values for common distributions in order to help choose distributions to fit data. A non-zero skewness would illustrate a lack of symmetry in a distribution of interest, while the kurtosis reveals the weight of the tails compared with those of the normal distribution (a normal distribution has a kurtosis value of 3).

Figure 3.1 shows a skewness-kurtosis plot for the distribution of logged change in Adélie penguin abundance. We estimated a skewness value of 1.023 and kurtosis

value of 6.135 for logged abundance transition data. To take into account the uncertainty of the estimated values of skewness and kurtosis, a non-parametric bootstrap procedure was also preformed, with reported values of skewness and kurtosis plotted along with those of the original observed data (Figure 3.1) [61]. The positive, non-zero value for skewness indicates that this distribution is right-skewed, and the kurtosis value reveals much heavier tails than the normal distribution. This gives valuable information to assist in the selection of candidate distributions with which to fit this data. It is worth noting that the right-skewed nature of this distribution implies that extreme events in this system are more often positive in nature, while the previous study from Anderson et al. [11] suggests that black-swans in animal populations usually manifest as population die-offs, not increases.

The large kurtosis value calculated for the empirical distribution for log-change in Adélie penguin abundance indicates a significantly heavier-tailed distribution compared to the normal used traditionally. In lieu of the normal distribution in Equation (1.4), a Student's t distribution may be more appropriate, allowing for varying levels of kurtosis greater than or equal to that of the normal distribution. In addition, we also consider other distributions with support for all real values and the potential for heavy tails, including the logistic, Cauchy, and Gumbel distributions [81, 84, 88]. Equivalently, we consider replacing the log-normal in Equation (1.5) with another distribution with support for non-negative numbers. In this case (see Figure B.1), a Gamma, Weibull, log-logistic, or Burr distribution should be considered as potential alternatives [12, 21, 69, 157].

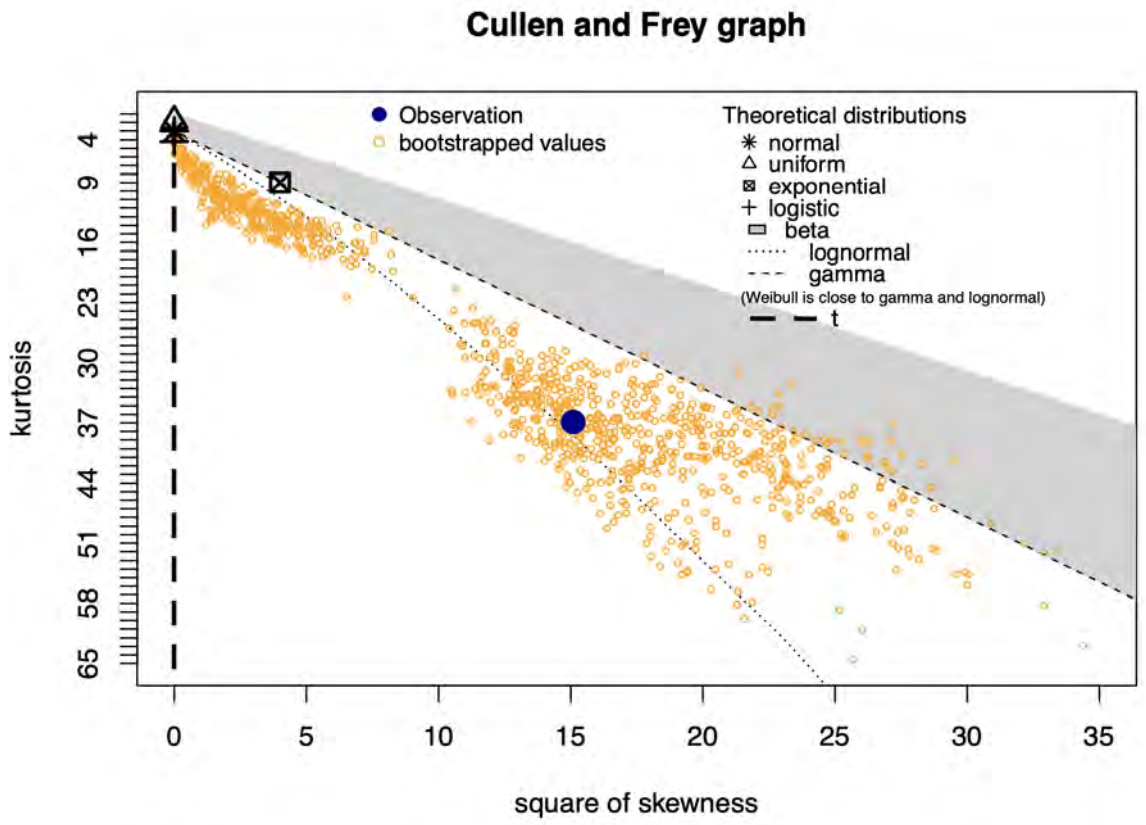

FIGURE 3.1: Skewness-kurtosis plot for empirical Adélie abundance data ($\log(\frac{N_{t+1}}{N_t})$), including values for 1000 bootstrap samples. Values are also given for several common distributions (the t distribution shown is the standard Student's t distribution).

3.2.2 Fitting candidate distributions

We next assess the fit of several candidate distributions for Adélie penguin abundance, replacing either the normal distribution on the log scale (Equation 1.4) or the log-normal distribution on the linear scale (Equation 1.5). The `fitdistrplus` package [61] for R was used to fit each of these parametric distributions to the empirical Adélie penguin data using maximum likelihood estimation (MLE). Goodness-of-fit statistics like the Anderson-Darling measure the distance between the empirical distribution and the fitted parametric distribution. We use the Anderson-Darling statistic in particular since it gives more weight to the tails of a distribution than other goodness-of-fit statistics like the Cramer-von Mises and Kolmogorov-Smirnov; for this reason, the

	AD	AIC	BIC	k
Normal	5.95	292.86	300.72	2
Student's t †	0.39	80.55	92.34	3
Logistic	2.27	133.52	141.38	2
Cauchy	2.10	125.28	133.14	2
Gumbel *	9.85	180.11	187.97	2

TABLE 3.1: The normal distribution and alternative distributions (Equation 1.4) fit to empirical Adélie abundance data ($\log(\frac{N_{t+1}}{N_t})$). Calculated Anderson-Darling statistic (AD), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values are shown as well as the number of parameters (k) for each distribution. †The nonstandard Student's t distribution, with mean μ, standard deviation σ, and degrees of freedom ν, is fit here (it is defined for use in R by the `ggdist` package [112]). *The Gumbel distribution is defined for use in R by the `VGAM` package [238]. All other distributions are defined in Base R. Parameter estimates for each fitted distribution are given in Table B.1.

Anderson-Darling statistic is often used in risk assessment and similar fields [56, 231]. However, since the Anderson-Darling statistic, like many others, does not consider the complexity of each model, we also calculate Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values to avoid over-fitting. AIC and BIC both penalize models based on the log-likelihood, discouraging the selection of those with more parameters to fit [30]. For all three of these goodness-of-fit measures, lower values indicate better fitting distributions.

On the logarithmic scale (Equation 1.4), several distributions out-performed the traditionally-used normal distribution (Table 3.1). While the logistic and Cauchy distributions both showed a better fit for logged Adélie penguin growth than the normal distribution, the Student's t distribution showed the best fit by far. The Student's t distribution used here is the nonstandard distribution, with mean μ and standard deviation σ as well as degrees of freedom ν (defined in R in the `ggdist` package [112]). This Student's t distribution was fit with an estimated degrees of freedom parameter of $\nu \approx 2.5$ (see Table B.1), accommodating the heavy tails of the empirical data (the normal distribution is roughly equivalent to a Student's t distribution with $\nu \geq 30$).

	AD	AIC	BIC	k
Log-normal	5.95	310.04	317.90	2
Gamma	8.43	625.82	633.68	2
Weibull	24.29	828.49	836.35	2
Log-logistic **	2.27	150.70	158.56	2
Burr **	2.27	132.31	144.10	3

TABLE 3.2: The log-normal distribution and alternative distributions (Equation 1.5) fit to empirical Adélie abundance data ($\frac{N_{t+1}}{N_t}$). Calculated Anderson-Darling statistic (AD), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values are shown as well as the number of parameters (k) for each distribution. **The log-logistic and Burr distributions are defined for use in R by the `actuar` package [66]. All other distributions are defined in Base R. Parameter estimates for each fitted distribution are given in Table B.2.

The log-normal distribution (Equation 1.5) is out-performed by both the log-logistic and Burr distributions when fitting Adélie penguin abundance on the linear scale (Table 3.2). The Burr distribution (Burr Type XII) can accommodate a wide range of values of skewness and kurtosis and is most commonly used to model household income (often referred to in this application as the Singh-Maddala distribution) [157, 213]. The log-logistic distribution is in fact a special case of the Burr distribution and is commonly used in survival analysis, such as mortality rate from cancer following treatment [21], and in economics to model wealth and income (referred to in this application as the Fisk distribution) [69].

Thus the traditional process model defined by Equation (1.4) (or, equivalently, Equation 1.5) is not the most accurate model for Adélie penguin abundance. The normal distribution traditionally used for abundance (Equation 1.4) underestimates the likelihood of extreme events and a heavy-tailed distribution, like the Student's t distribution, may be a more appropriate alternative. (Or, similarly, the log-normal distribution in Equation 1.5 may be replaced by a distribution like the Burr or log-logistic.) An accurate process model is critically important for modeling and forecasting population dynamics for any animal species and therefore, as this case study of Antarctic

penguins shows, heavy-tailed distributions should be considered as alternatives to the de facto process models which may vastly underestimate extreme events.

3.3 Fitting heavy-tailed distributions in JAGS

We next consider several of the alternative distributions for animal abundance investigated above and show how they can be modeled in a Bayesian framework for future modeling work for any animal species. In this analysis, rather than fitting particular observed data, we simulate data from each candidate distribution and use JAGS [176] to fit the distribution in question to that simulated data under a Bayesian framework. We first use the built-in JAGS function to define the distribution's likelihood, and then use both the zeros and ones tricks in turn [155]. In this way, we can compare the differences between using the built-in JAGS functions and using the more tunable, but less automatic, zeros or ones tricks. We also fit these distributions to the simulated data under a frequentist framework, using maximum likelihood estimation (MLE) as a baseline with which to compare the results of our Bayesian models.

We consider the alternative distributions for abundance given in Table 3.3, each with the potential for higher skewness and heavier tails than the normal distribution. We consider the logistic and Student's t distributions, to replace the normal distribution in Equation (1.4), and the Gamma and Weibull distributions, alternatives to the log-normal distribution in Equation (1.5). Though the log-logistic and Burr distributions show more promise for replacing the log-normal distribution in Equation (1.4) for Adélie penguins, we consider the Gamma and Weibull distributions because they are represented by built-in functions in JAGS. The use of the zeros and ones tricks here can act as a proof of concept for representing more uncommon distributions (specifically those without built-in JAGS functions) in the future. In addition, alternative distributions like the Gamma and Weibull may be more appropriate to replace the traditional log-normal distribution (Equation 1.4) for other animal species. For this

reason, we also include the logistic distribution as an alternative to the normal distribution in Equation (1.5) in addition to the Student's t.

3.3.1 Methods

Built-in JAGS functions

Table 3.3 gives the probability density function (PDF) for each candidate distributions, along with their parameterizations, as defined in both R and JAGS [176]. For several of these distributions, the parameterization used by the built-in JAGS functions differ from the (more conventional) parameterization used in R. Commonly, built-in functions in JAGS parameterize distributions using precision in lieu of the standard deviation. Of particular note is the case of the Student's t distribution, which is defined in Base R as the standard Student's t, having fixed mean 0 and standard deviation 1 (i.e. only parameterized by the degrees of freedom). The built-in function in JAGS for the Student's t distribution allows for nonstandard distributions, with non-zero mean and non-unity standard deviation, in addition to the varying kurtosis dictated by the degrees of freedom parameter [176]. The PDF of the nonstandard Student's t distribution has no simple closed-form expression. For this reason, we first consider the standard Student's t distribution (defined in Base R) to compare the efficacy of MLE, JAGS using the built-in function, JAGS using the zeros trick, and JAGS using the ones trick in fitting this distribution Next, we consider the nonstandard Student's t distribution (identified in Table 3.1 as a promising alternative to the normal distribution in Equation 1.4 and defined for use in R by the `ggdist` package [112]) to compare just the efficacy of MLE and JAGS using the built-in function in fitting this distribution.

Zeros trick

Consider the model $Y_i \sim G(\boldsymbol{\theta})$, where G is a distribution with parameters $\boldsymbol{\theta}$ and is *not* one of the distributions built into JAGS. Suppose the likelihood for the above model is

Distribution	PDF	R parameters	JAGS parameters
Logistic	$\frac{e^{-(y-\mu)/s}}{s(1+e^{-(y-\mu)/s})^2}$	`location` $= \mu$ `scale` $= s$	`mu` $= \mu$ `tau` $= 3/(s^2\pi^2)$
Student's t (Standard) †	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	(mean = 0, sd = 1) `df` $= \nu$	`mu` (mean) `tau` (precision) `k` $= \nu$
Student's t ‡	◇	`mu` $= \mu$ `sigma` $= \sqrt{\nu/(\nu-2)}$ `df` $= \nu$	`mu` $= \mu$ `tau` $= (\nu-2)/\nu$ `k` $= \nu$
Gamma	$\frac{\beta}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}$	`shape` $= \alpha$ `rate` $= \beta$	`r` $= \alpha$ `lambda` $= \beta$
Weibull	$\frac{k}{\lambda}\left(\frac{y}{\lambda}\right)^{k-1}e^{-(y/\lambda)^k}$	`shape` $= k$ `scale` $= \lambda$	`v` $= k$ `lambda` $= 1/\lambda$

TABLE 3.3: Probability density functions (PDFs) and parameterizations in both R and JAGS for the candidate distributions for abundance considered here. The logistic and Student's t distributions are alternatives to the normal distribution in Equation (1.4); the Gamma and Weibull distributions are alternatives to the log-normal distribution in Equation (1.5). †The Student's t distribution defined in Base R is the standard (mean 0, standard deviation 1) Student's t, while JAGS allows for nonstandard (non-central and/or non-standardized) Student's t distributions using its built-in function. ‡The nonstandard Student's t distribution, like the one defined by the built-in JAGS function, is defined for use in R by the `ggdist` package [112]. ◇The PDF of the nonstandard Student's t has no closed form, so the zeros and ones tricks cannot be used for this nonstandard distribution. All other distributions are defined in Base R. Built-in JAGS functions are described in the JAGS manual [176].

given as $p(y_i|\boldsymbol{\theta}) = L_i$. In order to model this in JAGS, suppose we have a set of observations $Z = 0$, each of which is assumed to be drawn from a Poisson(λ_i) distribution. If we wish to model $y_i \sim G(\boldsymbol{\theta})$ as above, we can set $\lambda_i = -\log(L_i)$ to obtain the desired likelihood distribution. Thus with $Z = 0$,

$$Z_i \sim \text{Poisson}(\lambda_i), \tag{3.2}$$

$$\lambda_i = -\log(L_i) \tag{3.3}$$

would result in the desired likelihood [121, 140, 155].

For example, to use the zeros trick to define the likelihood $p(y_i|\mu,\sigma) \propto \frac{1}{\sigma}e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}$ (the PDF of the normal distribution), the JAGS model would read as follows:

```
model {
  for (i in 1:length(y)){
    z[i] ~ dpois(lambda[i])
    lambda[i] <- log(sigma) + 0.5*pow((y[i] - mu)/sigma, 2)
  }
  sigma ~ dunif(0, 100)
  mu ~ dunif(-100, 100)
}
```

where `z` is loaded into the model as a vector of all zeros (in addition to the data `y`).

Ones trick

As an alternative to the zeros trick, consider again the model $Y_i \sim G(\boldsymbol{\theta})$ where G is a distribution that is not built into JAGS. Now suppose we have a set of observations $X = 1$, each of which is assumed to be drawn from a Bernoulli(p_i) distributions. To obtain the desired likelihood $L_i = p(y_i|\boldsymbol{\theta})$, we set $p_i = L_i$. Thus with $X = 1$,

$$X_i \sim \text{Bern}(p_i), \tag{3.4}$$

$$p_i = L_i \tag{3.5}$$

would result in the desired likelihood [121, 140, 155].

Below we include the JAGS code employing the ones trick for the same example shown in Section 3.3.1, with $p(y_i|\mu,\sigma) \propto \frac{1}{\sigma}e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}$ (the PDF of the normal distribution):

```
model {
  for (i in 1:length(y)){
    x[i] ~ dbern(p[i])
    p[i] <- pow(sigma, -1)*exp(-0.5*pow((y[i] - mu)/sigma, 2))
  }
  sigma ~ dunif(0, 100)
  mu ~ dunif(-100, 100)
}
```

where `x` is loaded into the model as a vector of all ones (in addition to the data `y`).

Procedure for each candidate distribution

For each candidate distribution G listed in Table 3.3, we simulate Y such that $Y \sim G(\boldsymbol{\theta})$ with known parameters $\boldsymbol{\theta}$. We then fit the original distribution G to the simulated data Y using four distinct methods in order to obtain parameter estimates for the parameters $\boldsymbol{\theta}$. First, the distribution is fit using MLE under a frequentist framework, employing the `fitdistrplus` package's `fitdist` function in R [61]. This will give a baseline for how well frequentist methods do in fitting these types of distributions. Next, the distribution is fit in a Bayesian framework using the built-in function provided in JAGS [176]. Lastly, the distribution will be fit in JAGS using both the zeros trick and the ones trick, as described above. This will allow for the comparison of not only the efficacy of fitting these distributions using Bayesian methods to that using frequentist methods, but also the relative performance of the built-in JAGS functions to the zeros and ones tricks and the relative performance of the zeros and ones tricks themselves.

All simulations are completed in R version 4.0.3 and JAGS version 4.3.0. Each JAGS model simulated 10,000 samples from the posterior after an initial burn-in period of 2,000 simulations.

3.3.2 Results

Parameter estimates for each candidate distribution are given for all four methods (MLE, JAGS using built-in functions, JAGS using the zeros trick, and JAGS using the ones trick) in Table 3.4 along with the true parameter values used to generate the simulated data of size $N = 1000$ (parameter values given for the JAGS models are posterior means). These results show how differently the four methods fit each distribution, both when parameterizations differ greatly between R and JAGS (namely for the logistic and Student's t distributions) and when parameterizations are more similar (Gamma distribution) or at least allow for more direct transformations (Weibull distribution). Trace plots and posteriors for each parameter in each JAGS model in Table 3.4 are given in Appendix B.2.

For most distributions, the built-in JAGS functions fit the simulated data similarly to the MLE procedure (Table 3.4). The exception is the logistic distribution, for which the model using the built-in JAGS function poorly estimates the shape parameter s. In this case, the zeros and ones tricks both perform virtually identically to MLE, estimating both parameters well. This may be on account of the fact that the built-in JAGS function model for the logistic distribution requires a potentially-troublesome transformation of the parameter s to the precision, $\tau = 3/(s^2\pi^2)$ (see Table 3.3); to add a complication to this transformation, JAGS does not define π, requiring a rounded estimate of π to be hard-coded when calculating precision. When using the zeros and ones tricks, the PDF of the logistic distribution is defined explicitly using the parameters μ and s, avoiding the built-in JAGS function's complicated parameterization. The precision must also be calculated for the model using the JAGS built-in function to fit the (nonstandard) Student's t distribution to simulated data. In this case, the estimation of the parameter σ also uses the precision, but the transformation ($\tau = (\nu - 2)/\nu$) is less complicated than that needed for the logistic distribution, and this JAGS model estimates σ as well as MLE does (Table 3.4). In all cases, the zeros and ones tricks

Distribution	Parameter	MLE	Built-in JAGS	Zeros trick	Ones trick
Logistic	$\mu = 1$	0.99	0.99	0.99	0.99
	$s = 2$	1.96	0.77	1.96	1.96
Student's t (Standard)	$\nu = 5$	5.08	5.15	5.25	5.26
Student's t	$\mu = 1$	0.98	0.98	$^\diamond$	$^\diamond$
	$\sigma = 0.5$	0.51	0.51		
	$\nu = 5$	5.07	5.24		
Gamma	$\alpha = 5$	4.92	4.92	4.93	4.92
	$\beta = 2$	1.93	1.93	1.94	1.93
Weibull	$k = 1.5$	1.52	1.52	1.52	1.52
	$\lambda = 1$	0.99	1.00	1.00	1.00

TABLE 3.4: Parameter values for each fitted distribution, as compared to true parameter values, for MLE (using the `fitdistrplus` package in R [61]), JAGS model with built-in function, JAGS model with zeros trick, and JAGS model with ones trick. Each model is fit to the same simulated dataset of size $N = 1000$. Parameter values for JAGS models are posterior means. Standard errors for each MLE parameter estimate are given in Table B.3; posteriors for each JAGS model parameter estimate are given in Appendix B.2.1. $^\diamond$The PDF of the nonstandard Student's t has no closed form, so the zeros and ones tricks cannot be used to model this nonstandard distribution.

fit each distribution to the corresponding data almost identically, with little difference between the parameter estimates. For the (standard) Student's t and Gamma distributions, the models using the ones and zeros tricks perform somewhat worse than MLE and the built-in JAGS function models.

Table 3.5 shows the results of a sensitivity analysis of all four methods to the amount of the data to be fit for the Weibull distribution (used here as a representative example of a candidate distribution), showing the difference in parameter estimates for decreasing sample sizes. There was no considerable difference between the performance of the various methods and decreasing dataset size, with all methods fitting the Weibull distribution increasingly poorly as sample size decreases below 500.

N	Parameter	MLE	Built-in JAGS	Zeros trick	Ones trick
1000	$k = 1.5$	1.52	1.52	1.52	1.52
	$\lambda = 1$	0.99	1.00	1.00	1.00
500	$k = 1.5$	1.53	1.53	1.53	1.53
	$\lambda = 1$	0.98	0.98	0.98	0.98
100	$k = 1.5$	1.41	1.40	1.40	1.40
	$\lambda = 1$	0.92	0.93	0.93	0.93
50	$k = 1.5$	1.60	1.57	1.58	1.57
	$\lambda = 1$	0.98	0.99	0.98	0.99

TABLE 3.5: Parameter values for fitted Weibull distributions, as compared to true parameter values, for MLE (using the `fitdistrplus` package in R [61]), JAGS model with built-in function, JAGS model with zeros trick, and JAGS model with ones trick. Parameter values for JAGS models are posterior means. Standard errors for each MLE parameter estimate are given in Table B.4; posteriors for each JAGS model parameter estimate are given in Figures B.12-B.14 and Appendix B.2.2. Models were run using $N = \{1000, 500, 100, 50\}$ data points.

3.4 Discussion

3.4.1 Adélie penguin case study

Recent work has found strong evidence for the occurrence of extreme black-swan events in many animal populations, most frequently in birds [11]. Here we used the Adélie penguin as a case study to identify heavy-tailed dynamics and determine the best-fitting distribution for abundance. We showed that the traditional process model defined by Equation (1.4) (or, equivalently, Equation 1.5) is not the most accurate model for Adélie penguin abundance, underestimating the likelihood of extreme events (see Tables 3.1 and 3.2). We identified several heavy-tailed distributions that may be more appropriate alternatives to the normal distribution in Equation (1.4) (or, similarly, the log-normal distribution in Equation 1.5). Namely, the Student's t distribution was the best fitting distribution for observed Adélie penguin abundance. Other distributions, including the logistic, Cauchy (replacing the normal in Equation 1.4),

log-logistic and Burr (replacing the log-normal in Equation 1.5), also out-preformed the classic process model for this Adélie case study.

There are many more distributions with the ability to accommodate heavier tails than that of the normal distribution (including some investigated here, see Tables 3.1 and 3.2) that may be the most fitting process model for other animal species. We urge practitioners to consider carefully a suite of models to determine the most accurate distribution for abundance, since, as we discuss next, many non-traditional distributions can be employed just as easily as the classic normal (i.e., log-normal on the linear scale) distribution.

3.4.2 Heavy-tailed models in JAGS: A proof of concept

We also illustrated how heavy-tailed distributions like the ones identified in the Aélie penguin case study can be modeled in a Bayesian framework using JAGS [177]. All candidate distributions were fit well using MLE, which serves as a good baseline with which to compare the three JAGS models (using the built-in JAGS functions, the zeros trick, and the ones trick; Table 3.4). The built-in JAGS functions often performed poorly when their parameterization differed greatly from the standard parameterization used in R—namely for the scale parameter (s) of the logistic distribution. For distributions with built-in JAGS functions that share a parameterization (or more direct transformations) with R, the built-in JAGS functions do a good job estimating parameter values, similar to the results from MLE. It is critical for one to understand the likelihood and parameterization of the built-in distributions they use, as parameterization differences between languages can have significant effects on analysis. In contrast, using the zeros or ones tricks in JAGS requires likelihoods to be defined explicitly, necessitating upfront an awareness of and appreciation for the distribution to be fit and allowing for the use of a desired parameterization. The models employing

the zeros and ones tricks in JAGS performed well in this analysis, avoiding parameterization confusion and the problems described above. Even for distributions whose built-in JAGS functions fit well, the zeros and ones tricks mostly perform similarly to MLE methods.

The zeros and ones tricks perform roughly equally to each other for all candidate distributions (Table 3.4). The major difference between the zeros and ones tricks is the scale on which the density function is defined. For the zeros trick, the (negative) logarithm is taken of the likelihood, so probabilities are on the log scale; for the ones trick, the likelihood is given on the original linear scale. For different likelihoods, one scale may lead to more numerical issues than the other. Other than these potential numerical issues, the zeros and ones tricks both perform well for the distributions tested here and, other than these potential numerical issues, are likely to perform similarly for other distributions. The fact that both methods are easy to implement in JAGS means that any distribution with a closed-form PDF can be defined using one of these 'tricks', including distributions that are not pre-defined by JAGS (or even those that are not defined by any R package).

Here, we only fit alternative distributions for abundance in JAGS where there existed built-in JAGS functions to define the distributions' likelihoods. This allows us to evaluate the efficacy of the ones and zeros tricks for use with more exotic distributions, particularly those that are not defined in JAGS as built-in functions, including the Burr, log-logistic, and Gumbel distributions (see Tables 3.1 and 3.2). Since these distributions are less commonly used than those such as the logistic and the Gamma, they do not have built-in JAGS functions and fitting them in JAGS would require the use of either of these tricks. This work serves as a proof of concept for the zeros and ones tricks as implemented in JAGS, allowing for a greater range of heavy-tailed (or otherwise more specific) distributions to be fit to observed animal abundance in future analyses and, importantly, the most appropriate process model to be employed in a given application.

3.4.3 Modeling alternative theories for extreme events

We have focused here on the black-swan theory of extreme events, aiming to represent growth rate by a single heavy-tailed distribution. However, the methods established here are not limited to the framework of this theory. An alternative theory suggests that, at least in some systems, extreme events are beyond the extrapolation of the power law or the use of heavy-tailed distributions for event generation. The study of these events refers to them not as black swans but as 'dragon kings', to emphasize that these extreme outliers ('kings') are a completely different animal ('dragons'), born of unique origins relative to typical events [205].

Under this theory, a process model may not be represented by one heavy-tailed distribution but instead by a mixture of multiple different distributions, representing the heterogeneous regimes of typical events and dragon-king events. These mixture models capture the bulk of the distribution with one model and either both of the tails with another distribution, or each of the tails separately with their own distinct parametric distributions. A simple mixture model might use one normal model (with a smaller standard deviation) for the bulk of the distribution and a second normal model (with a larger standard deviation) to capture the tails of the distribution. Another mixture model proposed by Mendes and Lopes [161] assumes the bulk of the distribution to be normal, with two separate generalized Pareto distribution (GPD) models used for each of the upper and lower tails. Such implementations of the dragon-king theory of extreme events may be represented using the zeros or ones tricks in JAGS, fitting mixture models of several distributions to animal abundance data under a Bayesian framework without relying on the built-in functionality of the software.

3.4.4 Accommodating skew

The Adélie penguin abundance data analyzed here as a case study for heavy-tailed abundance dynamics not only showed much heavier tails than the normal distribution

commonly used in models (e.g., [39, 104]), but also showed right skew (Figure 3.1); the normal distribution, symmetric about the mean, is unable to capture any such skew. The Student's t distribution, which was identified here (Table 3.1) as the best-fitting distribution for this empirical Adélie penguin data, unfortunately cannot accommodate any skew in the conventional parameterization (even the nonstandard parameterization used here), as it is also a symmetric distribution. To accurately include this observed skew in a process model, a skewed distribution should be employed. The skewed normal distribution [94] can be parameterized to include the level of right-skew seen in the Adélie data, but the tails of this distribution still lacks the support to correctly model extreme events. The generalized logistic (or, skew-logistic) [84] and skew-Cauchy [13] distributions are heavy-tailed distributions that may be able to more accurately replicate both the kurtosis and skewness seen in observed Adélie penguin data. In addition, the skewed Student's t distribution (also referred to as skewed generalized t distribution) [110] can accommodate both the skewness and kurtosis seen here; this distribution may be the most promising candidate, since the (symmetric) Student's t distribution showed the best fit to the empirical data out of the distributions considered (see Tables 3.1 and 3.2).

Chapter 4

Variability, skipped breeding, and heavy-tailed dynamics in an Antarctic seabird

4.1 Introduction

Variability of population abundance over time is an important facet of animal population dynamics [44, 59]. Large temporal variation in animal species hinders the ease with which time series for abundance can be used to accurately assess and predict population trends through time [23, 32, 191]. While large population fluctuations are seen across the animal kingdom [83], time series of abundance are particularly variable for colonially-breeding birds [11, 55]. Here, we study population abundance time series of the Adélie penguin (*Pygoscelis adeliae*), a well-studied colonially-breeding Antarctic seabird. The Adélie penguin shares life history traits with many long-lived colonially-breeding birds and serves here as a study system for a more general investigation into the causes and consequences of temporal population fluctuations in long-lived annually breeding species.

Heavy-tailed dynamics in animal populations and the associated extreme events [11] have significant consequences for population viability [55, 68]. Because of their disproportional effect on the systems in which they occur, these 'black-swan' events

are an important consideration in predicting the persistence and viability of animal populations [29, 149]. Additionally, these unexpected events may increase in frequency with increased human pressures on ecosystems [132]. For these reasons, developing an accurate model for time series dynamics requires an appropriate distribution for interannual variation in abundance or growth rates, with tails heavy enough to accommodate these rare events [2, 103].

Anderson et al. [11] studied extreme population-level events in over 600 animal species, finding evidence for the presence of black-swan events in about 4% of the animal populations studied and 7% of birds specifically. While some of these putative extreme events must be related to immigration/emigration rather than the dynamics of a closed population [239], it remains true that many time series display non-Gaussian dynamics that should be incorporated into ecological models. Here we consider the possibility of non-Gaussian heavy-tailed dynamics of Adélie penguin abundance for the first time.

4.1.1 Skipped breeding as a driver of stochasticity

The Adélie penguin is an important bellwether species long used to track the impacts of climate change and fishing on the Southern Ocean ecosystem, with considerable attention paid to the links between Adélie abundance and/or demography and environmental conditions. Though Adélie penguins are, barring extreme events [65], largely site faithful [4], their population dynamics are characterized by large interannual fluctuations [39, 111, 240]. While some environmental variables, notably sea ice concentration, have been convincingly linked to broad spatiotemporal patterns in Adélie trends [40, 104], interannual fluctuations in abundance at individual colonies make it challenging to deduce the causal driver of population changes through time [39]. Even more remarkably, time series of population abundance have low spatial coherence [198]; either Adélie penguins are responding to very fine scale environmental

conditions below the resolution of earth-observing satellites, or penguins at neighboring colonies are responding differently to the same ocean conditions. Either way, the links between the observed environment and Adélie population dynamics remain unclear, limiting our capacity to make precise forecasts of future abundance under climate change.

Like many long-lived species, the fitness of individual Adélie penguins is influenced by a large number of factors, the least understood of which is the decision to breed (or not) in any given year. Skipped breeding is common among many birds, especially among long-lived seabirds, and previous studies have demonstrated the importance of skipped breeding in understanding seabird population dynamics [54, 107, 108]. However, estimates for the rate at which Adélie penguins skip breeding are limited [16, 98] and the degree to which this drives population variability is unknown. Since population censuses of colonial seabirds are usually conducted by counting nests (or breeding pairs) present at the breeding colony, understanding the influence of skipped breeding is necessary if we are to use census data to determine population trends and forecast future population responses.

While skipped breeding has been estimated from flipper banding studies and has been explicitly incorporated into some age-structured models [16, 98], analyses of total penguin abundance through time [39, 40, 104, 143] typically include skipped breeding only implicitly as a component of model 'noise'. These simpler models leverage relatively widespread census data on the number of breeding pairs (equivalently, active nests) through time and not the more logistically challenging data obtained through flipper banding or related mark-recapture methods. As methods for population estimation using satellite imagery become more widely utilized [124, 143, 142], the volume of data available on total breeding abundance through time is poised to grow rapidly. It is in this context, in which abundance data is expanding in availability but is characterized by large and unexplained fluctuations that complicate our mechanistic understanding of population trends, that we sought to leverage the well-established

understanding of Adélie life history and demography to investigate how fluctuations in specific demographic rates (e.g., survival, reproduction, skipped breeding) manifest in fluctuations in total breeding abundance. This study was organized around the following research questions: 1) Which components of life history are actually reflected in time series of breeding abundance; 2) Are Adélie penguin dynamics accurately represented by Gaussian distributions for abundance, or are heavy-tailed distributions more appropriate for representing their variability structure; 3) Which demographic parameters are more likely to produce large interannual fluctuations in breeding abundance and, potentially, heavy-tailed dynamics; and, finally, 4) Do spatial patterns in the dynamics of Adélie breeding abundance correspond to patterns in environmental conditions?

4.2 Methods

4.2.1 Age-structured model

To represent the complicated life history of the Adélie penguin, which breeds annually during Antarctica's austral summer, we use an age-structured model that partitions the total population into classes based on age with varying vital rates [37]. Age-structured models have been used previously to capture the dynamics of many seabird species [97, 107, 109, 224]. Since seabirds like the Adélie penguin are long-lived and display delayed sexual maturity, age-structured models are an excellent tool for understanding the relationships between demographic rates and the abundance represented in annual surveys at the breeding colony.

We construct the age-structured model described in Figure 4.1, where reproductive success and survival to the next age class are represented as Binomial processes. The demographic parameters underpinning this age-structured model, which tracks

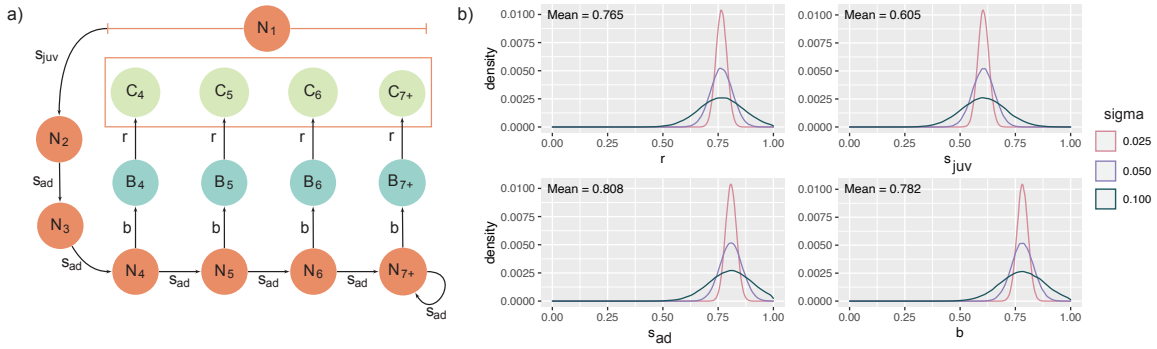

FIGURE 4.1: (**a**) Age-structured model diagram in which N_i is the number of (female) individuals in age class i (individuals in age class i are in their ith year of life, i.e. $i-1$ years old); B_i is the number of (female) individuals in N_i that attempted breeding in the current breeding season (only individuals in age classes $i = 4, 5, 6, 7+$ are eligible to breed; individuals may choose to skip breeding for the season, and not enter B_i with probability $1-b$); C_i is the number of chicks (both male and female) successfully produced by breeding class B_i. Each individual in B_i lays two eggs with her partner; those eggs hatch and survive that breeding season each with probability r. N_1 is the total number of chicks (both female and male) that survived through the breeding season in which they were born and is equal to $C_4 + C_5 + C_6 + C_{7+}$. The next breeding season, the number of (female) individuals in N_2 is drawn from a Binomial distribution with $n = N_1/2$ and probability s_{juv}. Breeding abundance, here often referred to as abundance, is equal to the sum $B_4 + B_5 + B_6 + B_{7+}$. (**b**) Density plots for demographic parameters (top: r, reproductive success, and s_{juv}, juvenile survival; bottom: s_{ad}, adult survival, and b, breeding propensity). Parameter values for each year are drawn from truncated normal distributions centered about the mean values for each parameter (equal to the posterior medians from ABC parameter estimation), shown here with standard deviations $\sigma = 0.025, 0.05$, and 0.1 (σ values used in simulations range from $0.001 - 0.1$).

female penguins only, are reproductive success r, juvenile survival s_{juv}, adult survival s_{ad}, and breeding propensity b. The last of these key demographic rates, breeding propensity, is the probability that a sexually mature penguin will choose to breed in any given season. In this way, we allow for some number of individuals to skip breeding each year (i.e., with probability $1-b$), a life history trait of particular interest for our analysis. Adélie penguins overwhelmingly lay two eggs in each clutch and we assume each nest is comprised of two laid eggs, each of which has probability r of successfully surviving to crèche (here defined as the point at which chicks are no

longer closely incubated on the nest). The model easily accommodates age-specific reproductive rates r_i, survival rates s_i, and breeding probabilities b_i, but other than distinguishing between adult and juvenile survival, we present a simplified model in which demographic rates are assumed the same across age classes. This reduced complexity allows us to focus on the links between demographic stochasticity and the total number of breeding pairs without loss of generality.

4.2.2 Parameterizing age-structured simulations

In order to simulate time series using demographic rates that reflect realistic scenarios, we use Approximate Bayesian Computation (ABC) to estimate the true demographic rates reflected in existing time series of Adélie penguin abundance through time at Cape Crozier, one of the most well-studied Adélie colonies in Antarctica [64, 65, 145]. ABC methods do not require the evaluation of the likelihood function and thus can be used to estimate parameter values in complex models (see [138] for a review of ABC methods and their use in population models). We use the age-structured model described above to estimate the values of the four demographic rates, reproductive success (r), juvenile survival (s_{juv}), adult survival (s_{ad}), and breeding propensity (b) by simulating time series using randomly drawn values for each of the four parameters and then accepting or rejecting those time series (and thus their parameter sets) based on their similarity (as described below) to observed time series of Adélie penguin breeding abundance. In this way, we infer the underlying demographic rates through their ability to generate dynamics commensurate with observation.

The initial population for the simulated time series is set to the actual Adélie nest count from Cape Crozier in 1985 [144, 145], chosen since it is a large site for which there is a nearly complete 30+ year time series and a history of mark-recapture studies [64, 65, 145]. For each simulation, parameter values for the initial year $t = 1$ (θ_1) are drawn

from Beta-distributed priors moment-matched to published estimates [64, 65, 97]. After the initial year, demographic parameters for year $t > 1$ are drawn from truncated $(0, 1)$ normal distributions ($\sigma = 0.025$) centered around the initial year's parameter value for that time series (i.e., $\theta_{t+1} \sim TN(\mu = \theta_1, \sigma = 0.025, 0, 1)$). Simulated time series are accepted or rejected based on mean absolute percentage error (MAPE) from the full observed time series from Cape Crozier (data from 1982-2018, [144]). For computational tractability we aimed for a 0.1 acceptance rate and set the MAPE threshold accordingly. A total of $30,000$ simulated time series yielded 2772 accepted parameter sets. The median of the accepted values for each demographic parameter was used in the simulation study described below. Additional details on the estimation of demographic rates are included in Appendix C.1. Note that our extraction of demographic rates from the published time series simply ensures that our simulated time series are realistic, but the precise values used for average survival, reproduction, and breeding propensity are largely inconsequential to our study of how variation in these factors propagates to the dynamics of total breeding pairs.

4.2.3 Age-structured simulations

We simulate 75-year time series of breeding abundance (defined as the total number of breeding pairs, or $B_{\text{total}} = B_4 + B_5 + B_6 + B_{7+}$) via the age-structured model illustrated in Figure 4.1, seeded with identical initial populations. The value of each of the four demographic parameters of the model r, s_{juv}, s_{ad} and b is centered on a constant mean value equal to the posterior median from the ABC described above; these rates are similar to those estimated by previous studies [64, 65, 97]. In each year t, values for each parameter are drawn from a normal distribution truncated $(0, 1)$ with standard deviation σ (see Figure 4.1b). While true Adélie penguin demographic rates may be correlated with one another, we assume for simplicity that each demographic parameter is realized from the truncated normal distribution independently each year. The

number of individuals in each age class in year t ($N_{i,t}$) is calculated using the previous year's totals ($N_{i,t-1}$) and the current year's adult survival ($s_{ad,t}$). Numbers of breeders and chicks in year t ($C_{i,t}$ and $B_{i,t}$) are calculated using the current year's totals ($B_{i,t}$ and $N_{i,t}$, respectively) and the rest of the current year's demographic parameters (r_t, $s_{juv,t}$, and b_t). We simulate $n = 10$ time series for each value of σ in $[0.001, 0.1]$, here broken into $m = 15$ steps, for each parameter. For each step in this $15 \times 15 \times 15 \times 15$ 4-dimensional space, 10 separate time series iterates are simulated. We also calculate observed growth in abundance for each year of each iteration as the logged change in total breeding abundance, $\text{ObsGrowth}_t = \log((\sum_{i=4}^{7+} B_{i,t})/(\sum_{i=4}^{7+} B_{i,t-1}))$. MAPE is used as a metric for how closely growth in breeding abundance matches the true growth in abundance, $\text{TrueGrowth}_t = \log((\sum_{i=4}^{7+} N_{i,t})/(\sum_{i=4}^{7+} N_{i,t-1}))$. Note that, since the age-structured model is simulated using Binomial draws, the simulations are stochastic.

The described stimulation study was repeated for non-Gaussian, heavy-tailed inputs. Each year, parameter values are realized using a Student's t distribution centered on the same constant mean values as the corresponding truncated normal distributions described above, with standard deviation σ and degrees of freedom parameter $\nu = 3$.

4.2.4 Detecting heavy tails in observed abundance time series

We used a Bayesian state-space model to estimate annual Adélie nest abundances for all known Adélie breeding sites from 1970 - 2020, grouped according to Antarctic Conservation Biogeographic Regions (ACBRs, see [219]). We used an exponential growth rate model adapted from [39] that included observation error (uncertainty in the number of true nests counted in each year) and process error (stochastic variability in the growth rate r) but, unlike [39], included no environmental covariates modulating the average growth rate at each site. In addition, we replaced the Gaussian distribution for true nest abundance in the process model (see equation 5 in Supplementary Data

1 of [39]) with a shifted (mean $\not\equiv 0$) and scaled (standard deviation $\not\equiv 1$) Student's t distribution. Thus, we model the true (latent) Adélie nest abundance $z_{i,t}$ at the i^{th} breeding site located in ACBR $R[i]$ in the year t as:

$$\log(z_{i,t}) \sim \text{student-t}(\mu_{i,t} = \log(z_{i,t-1}e^{\gamma_{R[i]}}), \sigma^2_{R[i]}, \nu_{R[i]}), \tag{4.1}$$

where the standard deviation (σ) and degrees-of-freedom (ν) of the Student's t distribution are estimated for each ACBR, and $\gamma_{R[i]}$ is the region-specific mean growth rate. Each of these parameter values ($\sigma_{R[i]}$, $\nu_{R[i]}$, and $\gamma_{R[i]}$) is shared across each site in an ACBR. The parameter ν of the Student's t distribution dictates the heaviness of the tails of the distribution: the smaller the value of ν, the heavier the tails of the distribution; with $\nu \geq 30$, a Student's t distribution is indistinguishable from the Gaussian distribution with the same mean and standard deviation [128]. Using the estimated values for ν, we classified each of the ACBR time series dynamics as either "heavy-tailed" ($\nu < 30$) or Gaussian ($\nu \geq 30$). We do not include any site or season effects. A more detailed model description is included in Appendix C.2.

4.2.5 Krill/Environmental model

We used output from the Community Earth System Model version 2 (CESM2, see [58]) to investigate spatial patterns in food resource variability for Adélie penguins. CESM2 is a state-of-the-art Earth system model with a nominal 1° resolution. The ocean model in CESM2 is the Parallel Ocean Program (POP), a level-coordinate ocean general circulation model with 60 vertical levels. The marine ecosystem in POP is simulated using the Marine Biogeochemical Library (MARBL). A thorough description of the ocean ecosystem model in CESM2 is documented in [137]. The flexible structure of MARBL facilitates the addition of multiple phytoplankton and zooplankton functional types. A detailed description of the MARBL configuration used here is included in Appendix C.3.

The CESM2 simulation we used in this study was an ocean-sea-ice-only simulation (i.e., only the ocean and sea ice components of CESM were active) forced by the Japanese Reanalysis (JRA) atmospheric dataset [116]. One JRA "cycle" is 61 years long, representing the observed historical atmospheric forcing over the period 1958 to 2018. We performed two JRA cycles with CESM2 (122 simulation years total), and only analyzed the second cycle when the model was "spun up". This simulation may be considered a global historical reconstruction of the physical ocean and marine ecosystem over the period of atmospheric forcing, 1958 to 2018.

The diet of Adélie penguins primarily consists of krill and small fish [223]. To understand how food resources for Adélie penguins may vary year-to-year, we computed interannual anomalies (detrended annual means) of various ocean ecosystem variables over the 1958 to 2018 period. The macrozooplankton group in CESM is representative of krill and other large crustacean zooplankton and therefore could be consumed directly by Adélie penguins. The mesozooplankton group, however, represents smaller zooplankton crustaceans, such as copepods, and could serve as a food source for fish, which could, in turn, be consumed by penguins; this group, therefore, can be considered an indirect food source for Adélie penguins. In order to simplify the analysis and capture of full potential of food resources for Adélie penguins, we grouped the small microzooplankton and microzooplankton biomass together as "small zooplankton" and mesozooplankton and macrozooplankton biomass together as "large zooplankton", with the large zooplankton group being representative of food resources for Adélie penguins. We depth-integrated zooplankton biomass fields over the top 150m of the ocean. Finally, we computed the standard deviation of interannual anomalies on a gridcell-by-gridcell basis of depth-integrated biomass of large and small zooplankton, as well as other potential sources of variability, such as net primary production (NPP) and sea ice coverage, and examined geographic patterns of the standard deviation of these variables.

4.3 Results

The age-structured model framework and the resulting relationship between changes in demography and the observed breeding abundance are illustrated in Figure 4.2. In particular, year t breeding propensity (b) affects year t breeding abundance, but all other effects are lagged. Changes in adult survival (s_{ad}) in year t are not seen in abundance until year $t+1$, while changes in juvenile survival (s_{juv}) and reproductive success (r) are not seen until years $t+3$ and $t+4$, respectively. Breeding propensity has the most complicated effect on observed breeding abundance. A single change in breeding propensity in year t is seen twice in breeding abundance: the first is immediately in year t, since changing the proportion of individuals who breed in a given year directly affects the number of birds present at the colony for observation. The second effect on abundance of changing breeding propensity occurs in year $t+5$, a full model cycle later. This signal is much weaker, as it has been dampened over time by the other demographic parameters. Though these relationships are direct translations of the model description, two important observations seen here will shape the remainder of the results (Figure 4.2): (1) extreme changes in breeding propensity cause large, immediate fluctuations in observed breeding abundance that obscure true trends but generate only small changes in long-term abundance; and (2) extreme events in adult survival have the largest long-lasting impact on breeding abundance.

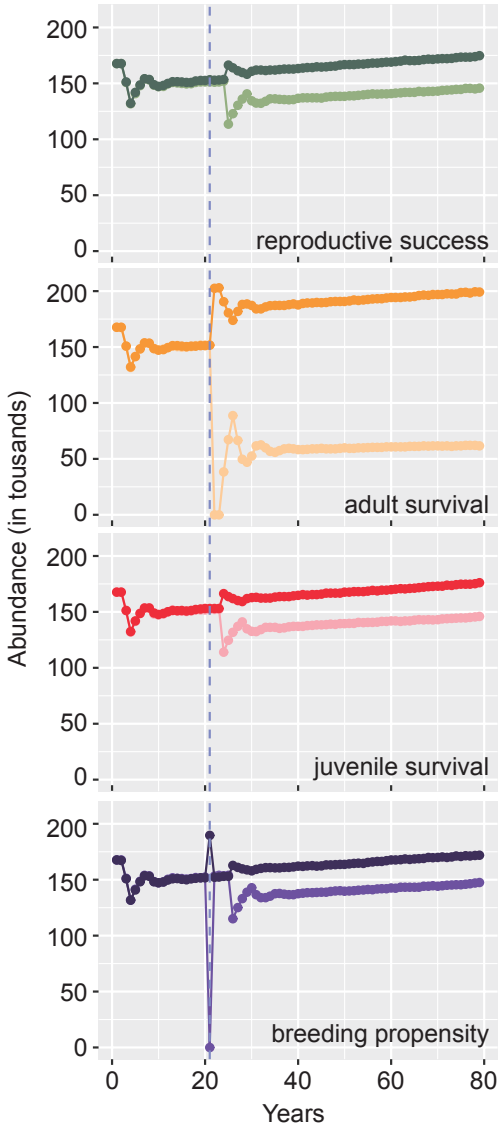

FIGURE 4.2: Time series for abundance with constant demographic parameters except in year $t = 20$, when a positive (darker shade) or negative (lighter shade) extreme event occurs for one of the parameters. Base values for parameters are $r = 0.75$, $s_{juv} = 0.75$, $s_{ad} = 0.75$, and $b = 0.80$. When an extreme event occurs for one of the four demographic parameters, its value is set to 1 (positive extreme events) or 0 (negative extreme events).

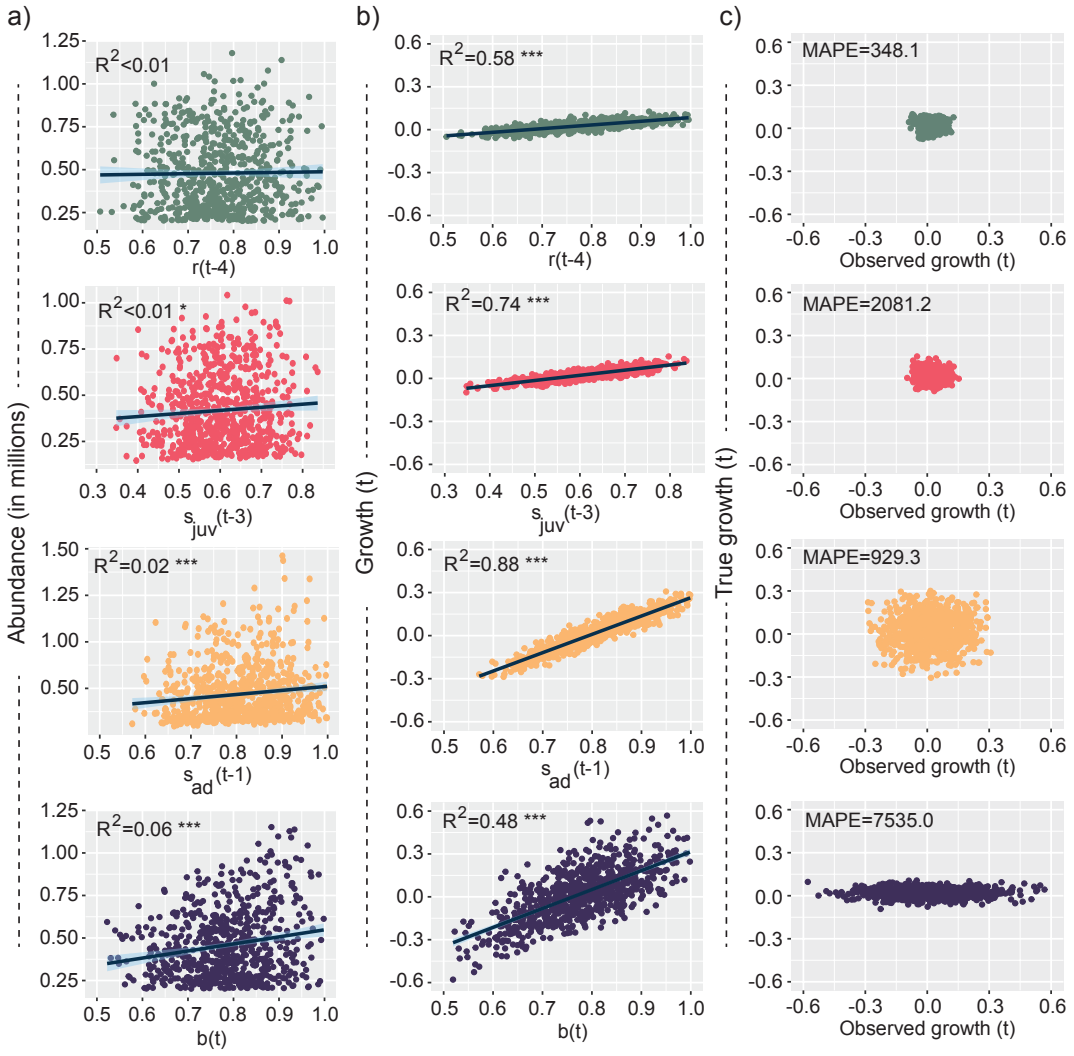

FIGURE 4.3: Scatter plots of (a) breeding abundance and (b) growth in breeding abundance vs. each of the four demographic parameters with various lags. Lags are as follows: 4 years for r, 3 years for s_{juv}, 1 year for s_{ad}, and 0 years for b. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$; * indicates a significant correlation with $p < 0.05$. (c) Scatter plots of true growth in breeding abundance vs. observed growth in breeding abundance. In each plot, the standard deviation of one parameter (from top: reproductive success r, juvenile survival s_{juv}, adult survival s_{ad}, and breeding propensity b) is high (0.1), while the standard deviations of the other three parameters are low (0.001). Mean absolute percentage error (MAPE) values are shown for each case.

4.3.1 Breeding propensity and adult survival drive observed abundance

Each demographic parameter affects breeding abundance on a different time scale (Figs 4.2, 4.3a). Considering each demographic parameter with the lags identified in Figure 4.2, breeding abundance is most strongly correlated with breeding propensity in year t (positive correlation, $R^2 = 0.0561$, $p < .001$; Figure 4.3a). No correlation is seen between breeding abundance and the other three demographic parameters (reproductive success, juvenile survival, and adult survival) in the current year t (see Figure C.8). However, breeding abundance in year t was positively correlated with adult survival in the previous year $t - 1$ ($R^2 = 0.0158$, $p < .001$; Figure 4.3a; see also Figure C.9) and, to a lesser extent, positively correlated with the juvenile survival from year $t - 3$ ($R^2 = 0.0074$, $p < .05$; Figure 4.3a; see also Figure C.11). Breeding abundance in year t was also negatively correlated with adult survival in year $t - 2$ (see Figure C.10). When all parameters vary maximally (i.e. the standard deviation of each parameter is 0.1), a positive correlation between abundance and that year's breeding propensity is still evident ($R^2 = 0.0053$, $p < .05$; see Figure C.20), but it is dominated by the correlation with adult survival from the previous year $t - 1$ ($R^2 = 0.02$, $p < .001$).

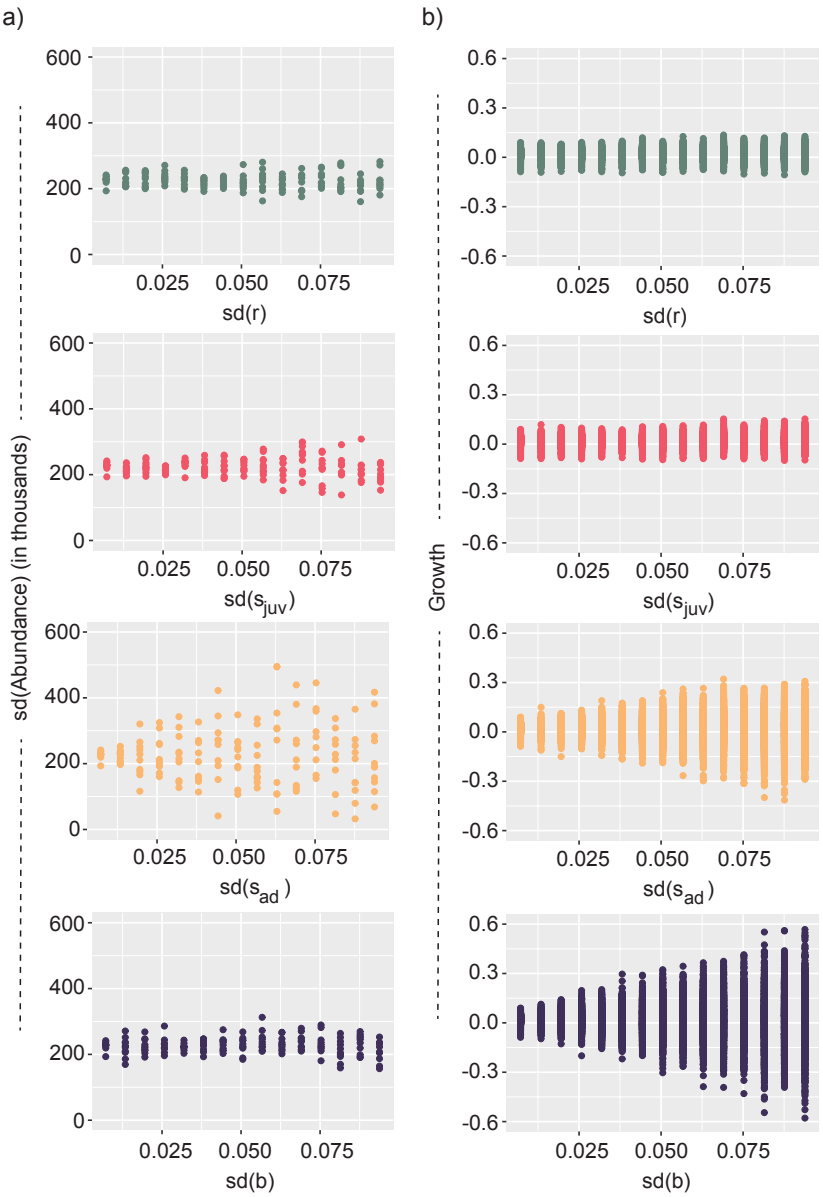

FIGURE 4.4: (a) Scatter plots of standard deviation in breeding abundance vs. standard deviation in each of the four demographic parameters. In each plot, the standard deviation of the parameter of interest increases from 0.001 to 0.1 along the x-axis (with 10 iterates per value of σ, one for each identical stochastic run of the simulation study), while the standard deviations of the other three parameters are constant at 0.001. (b) Scatter plots of growth in breeding abundance vs. standard deviation in each of the four demographic parameters. In each plot, the standard deviation of the parameter of interest increases from 0.001 to 0.1 along the x-axis (with a data point for each consecutive pair of years in the simulated time series for each of 10 iterations of the simulation for each value of σ), while the standard deviations of the other three parameters are constant at 0.001.

4.3.2 High variability in breeding abundance decouples observed and true abundance

All four demographic parameters have significant positive correlations with the growth of breeding abundance in year t when the above lags are considered (Figure 4.3b; additional results in Appendix C.4). MAPE is highest for the case where breeding propensity is the only parameter varying maximally (Figure 4.3c). Thus, when breeding propensity is highly variable, observed growth rates (i.e., growth in breeding abundance) is least predictive of true growth in abundance.

Using ordinary least squares regression, there are no significant relationships between the standard deviation of abundance and the standard deviation of any of the demographic parameters (Figure 4.4a); however, errors are shown to be heteroskedastic in some cases. With growing variation in adult survival and, to a lesser extent, juvenile survival, the spread of the standard deviation in abundance (over 10 simulations) increases. However, the value of sd(Abundance) remains centered at about 200,000 over all simulations. Heteroskedasticity is also observed for the growth in breeding abundance against variability in adult survival and breeding propensity (Figure 4.4b), with the spread of growth increasing most for high variability in breeding propensity. Additionally, all four demographic parameters have significant positive linear relationships between their variability and the variability of growth in breeding abundance (see Figure C.21). Variability in breeding propensity had the largest impact on variability in growth, followed by variability in adult survival.

4.3.3 Spatial patterns in heavy-tailed abundance dynamics correspond to regions of high environmental variability

Adélie colonies in the Northwest Antarctic Peninsula (NW AP, ACBR 3; Figure 4.5) were determined to have solidly heavy-tailed dynamics (posterior mean of $\bar{\nu}_{R_3} = 3.1$; see Appendix C.2). The South Orkney Islands (ACBR 2) were also classified as having

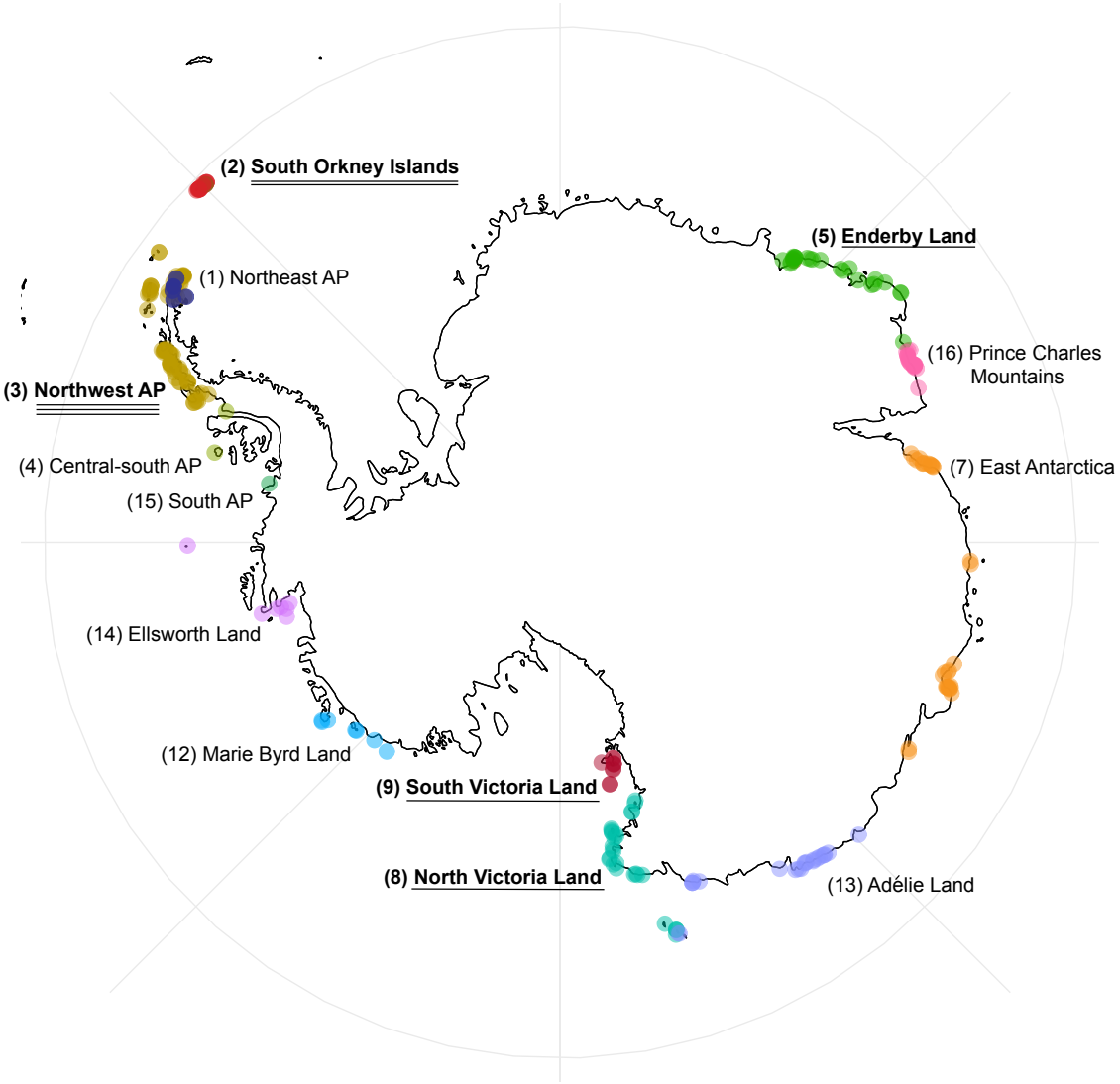

FIGURE 4.5: Antarctic Conservation Biogeographic Regions [219] labeled according to Adélie penguin population dynamics, as determined by values of the degrees of freedom parameter (ν) of the Student's t distribution fitted to nest abundance in our Bayesian state-space model. Region names are underlined according to degree of heavy-tailed dynamics identified: very heavy-tailed ($\nu \leq 5$; triple underline), moderately heavy-tailed ($5 < \nu \leq 15$; double underline), or slightly heavy-tailed ($15 < \nu < 30$; single underline). Regions with no underline did not have enough data to gain low enough prior-posterior overlap to present results. Full results are given in the Appendix C.2.

heavy-tailed dynamics, but to a lesser extent ($\bar{\nu}_{R_2} = 13.2$). North Victoria Land (ACBR 8), South Victoria Land (ACBR 9), and Enderby Land (ACBR 5) were each determined to have dynamics that are closer to Gaussian ($\bar{\nu}_{R_8} = 21.3$, $\bar{\nu}_{R_9} = 26.6$, $\bar{\nu}_{R_5} = 25.1$), but still below the threshold of $\nu \geq 30$ for identically Gaussian dynamics. Because

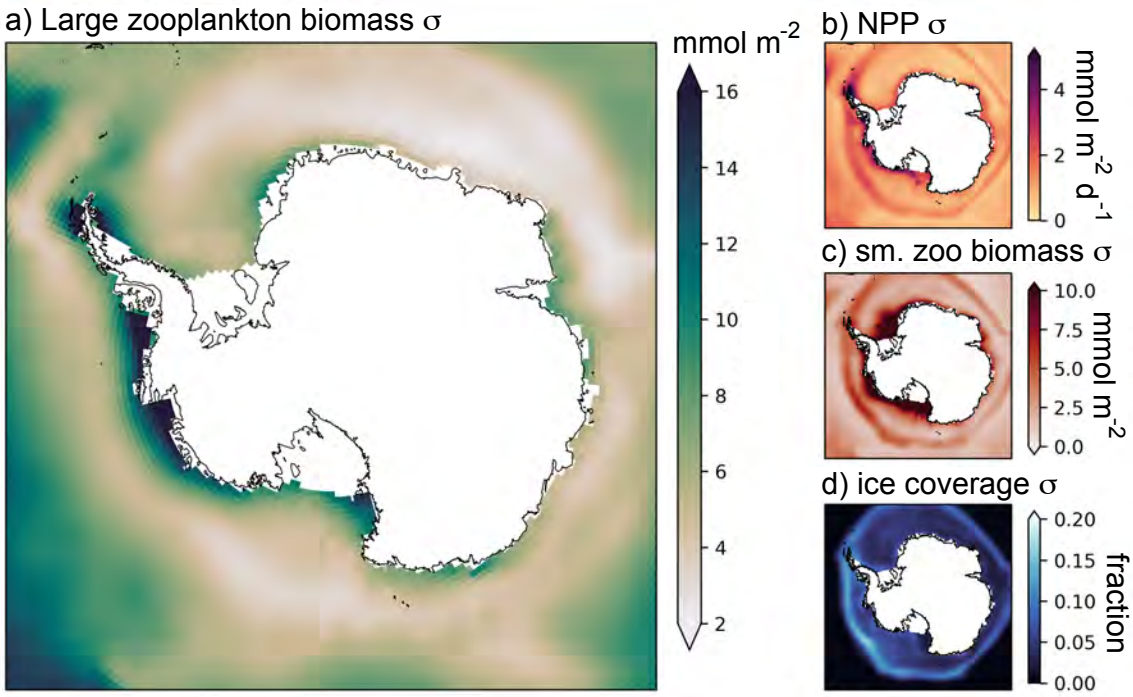

FIGURE 4.6: Interannual variability of ecosystem variables from CESM2 over the simulation period 1958 to 2018 for the Antarctic region. Panel (a) shows the standard deviation (σ) of modeled depth-integrated large zooplankton biomass. Panels (b), (c), and (d) show standard deviation (σ) of depth-integrated net primary productivity (NPP), depth-integrated small zooplankton biomass, and fractional sea ice coverage, respectively.

of differences in data density, we have the highest degree of confidence in the results for the NW AP and South Orkney Islands and less confidence in the results for North and South Victoria Land and Enderby Land. There was insufficient data within the other ACBRs (uncolored in Figure 4.5) for the model to estimate the parameter ν (as determined by the overlap of the posterior distribution with the prior and measures for convergence) and these results are included in Appendix C.2 but omitted from further consideration.

Heavy-tailed dynamics in Adélie abundance correspond with regions of high environmental variability (Figure 4.6). In particular, the regions identified as having heavy-tailed dynamics—the NW AP, South Orkney Islands, and North and South Victoria Land—all experience high variation in large zooplankton biomass, suggesting

that variability in food resources could be contributing to variability in Adélie abundance (Figure 4.6a). Additionally, the region which displayed the most significantly heavy-tailed Adélie abundance dynamics, the NW AP, also experiences the highest levels of variability in large zooplankton biomass and NPP across the Southern Ocean (Figure 4.6a,b); this is in contrast to small zooplankton biomass variability, which is relatively low in the NW AP (Figure 4.6c). Variability in sea ice coverage (Figure 4.6d) may also influence primary production and krill abundance, especially in areas like the NW AP where sea ice variability is high (e.g., see [189]). While disentangling the links between zooplankton and sea ice variability is beyond the scope of this paper, it is worth pointing out that variability in sea ice coverage in areas such as the NW AP could also be contributing to heavy-tailed Adélie abundance dynamics in this region.

4.3.4 Highly variable adult survival drives heavy-tailed abundance dynamics

In the simulation study, distributions for abundance resulting from simulations with combinations of highly variable demographic parameters were fit with Student's t distributions. We found a significantly higher ($p < .001$) probability of obtaining time series with heavy-tailed dynamics with highly variable adult survival ($\sigma = 0.1$) than with minimally variable adult survival ($\sigma = 0.001$), regardless of the variation in the other demographic parameters (see Figure C.23). In fact, the only inputs to the system that resulted in a probability of heavy-tailed dynamics any higher than 0.2 had highly variable adult survival rates; in these cases, the probability of heavy-tailed abundance dynamics ranged from 0.3 - 0.8 (mean = 0.4). Adult survival is the only vital rate in which an increase in probability of heavy-tailed dynamics is seen with increasing variability (see Figure C.22).

Results of the simulation study using non-Gaussian heavy-tailed inputs for demography generally followed those obtained using Gaussian variation: abundance in

year t is correlated with breeding propensity in year t, adult survival in year $t-1$, juvenile survival in year $t-3$. In this case, abundance in year t was also correlated with reproductive success in year $t-4$ (see Figure C.25). Notably, correlations between variation in abundance and variation in reproductive success, adult survival, and breeding propensity were found when demography varies according to a Student's t distribution (Figure C.27). Surprisingly, these correlations are all negative, meaning increased heavy-tailed variation in these demographic parameters leads to decreased variation in breeding abundance. Further, we found that variation in abundance is significantly lower for high variation Student's t inputs than it is for low variation Student's t inputs (Figure C.28). In the context of these results, it is important to note that our model does not include density dependence. Additionally, when parameters are drawn from a Student's t distribution, there is still a significant increase in the probability of heavy-tailed abundance dynamics seen with increased variability of adult survival (Figure C.29), as was observed for Gaussian inputs. Interestingly, we found significantly less variation in breeding abundance in time series resulting from heavy-tailed inputs than those with Gaussian inputs (see Figure C.30).

4.4 Discussion

Since population censuses of colonial seabirds like the Adélie penguin are usually conducted by counting nests (or breeding pairs) present at breeding colonies, skipped breeding is fundamental to our understanding of how census data should be used to estimate trends and forecast population responses. We have shown that breeding propensity drives observed breeding abundance (Figure 4.3a) and, importantly, fluctuations in growth of breeding abundance (Figure 4.4b) more than any other vital rate. Therefore skipped breeding interferes with our ability to link breeding abundance fluctuations with any other component of Adélie penguin life history. Predictably,

high variability in breeding propensity also caused the largest decoupling between true growth in abundance and observed growth (Figure 4.3c).

While some have argued that skipped breeding among Adélie penguins is too infrequent to be important [16], other studies have indicated that many individuals may choose to defer breeding in some years [4, 151]. Given the direct link between skipped breeding and abundance in a given year, and the capacity of skipped breeding to obfuscate other important links between environmental conditions and Adélie penguin dynamics, this aspect of Adélie life history is understudied. It is worth emphasizing that uniformly high rates of skipped breeding are not as problematic to our interpretation of census data as variable skipped breeding, since the long-run average population at the breeding colony would remain indexed to the total population of breeding-age adults. High interannual variability in breeding propensity, even if skipped breeding is low on average, may explain the difficulty of correlating fluctuations in abundance to environmental conditions [39], the short forecast horizon for Adélie dynamics [101], and the lack of spatial and temporal transferability of population trends and their environmental drivers [198].

The conservation implications of this finding are considerable, not just for Adélie penguins but potentially for other species whose life history strategy includes skipped breeding, such as many seabirds. In the Antarctic, the Adélie penguin is one of several species regularly monitored by the Commission for the Conservation of Antarctic Marine Living Resources Ecosystem Monitoring Program (CEMP), whose stated objectives are to "monitor the key life-history parameters of selected dependent species" (e.g., the Adélie penguin) as a means by which to track populations of target species like Antarctic krill, on which much of the Antarctic food web depends [49]. Specifically, "[s]uitable 'indicator species' should show measurable responses to changes in the availability of [krill], for example in changes in population size" [49]. Our results here provide one explanation for why fluctuations in abundance may be so difficult to link to environmental conditions and imply that other metrics tracked by CEMP (such

as foraging behavior) may be more valuable for monitoring the impact of krill fishing on krill predators such as the Adélie penguin.

The effects of varying each demographic rate in isolation on breeding abundance were not consistent with the effects resulting from multiple rates varying simultaneously. For example, when all demographic rates are varying maximally (Figure C.20), observed breeding abundance had the strongest correlation with adult survival (lagged one year). Further up the life cycle, a positive correlation between abundance and breeding propensity (in the current year) was present, but it was less strong than the correlation with adult survival. This suggests that the effect of variable breeding propensity is less important when multiple vital rates are highly variable, though it will still serve to obscure true abundance and growth trends observed from breeding abundance.

The longevity of the effect of variability of demography on breeding abundance also depended on the type of the demographic rate. For example, while variability in breeding propensity causes large immediate interannual fluctuations in observed breeding abundance, extreme skipped breeding events generate only small long-term changes in abundance (Figure 4.2). Extreme events in adult survival, however, cause long-term changes in both observed and true abundance. It is important to highlight the fact that breeding propensity and adult survival are independent in this age-structured model, though in reality life history trade-offs and density dependence (not included here) likely drive a correlation between them over the long term.

While breeding abundance was correlated with breeding propensity and adult survival with lags dictated by the structure of the age-structured model (Figure 4.3a, lags illustrated in Figure 4.2), there was no correlation between breeding abundance and reproductive success (4-year lag, Figure 4.3a), and only a small correlation between breeding abundance and juvenile survival (3-year lag, Figure 4.3a). Further, we found no correlation between breeding abundance in year $t-5$ and breeding propensity (see Figure C.13), while Figure 4.2 shows a clear effect of an extreme breeding propensity

event on abundance after 5 years. Thus the "last" (i.e., furthest down the life cycle) correlation between breeding abundance and demography was with juvenile survival in year $t-3$. Perhaps since the value of s_{juv} is relatively low [64, 65, 97], it acts to erase fluctuations from earlier stages in the life cycle. Thus lagged signals from reproductive success, breeding propensity and—to a lesser extent—juvenile survival itself, do not propagate to abundance. However, fluctuations from adult survival (in year $t-1$) and breeding propensity (in year t) emerge more strongly in abundance since they affect breeding abundance after juvenile survival in the life cycle. It is important to note that, even where we find correlations between breeding abundance and demography, explained variance (R^2) is low, emphasizing how difficult it is to link observed breeding abundance at the colony to demographic life-history parameters.

4.4.1 Heavy-tailed dynamics

Extreme fluctuations in breeding abundance, either caused by breeding propensity or adult survival, do not necessarily follow a Gaussian distribution. Here we showed that many regions of Antarctica are characterized by a distribution for abundance with heavier tails than befit a Gaussian distribution—with the most significant heavy-tailed dynamics occurring in the Northwest Antarctic Peninsula (Figure 4.5). We also found that the presence of heavy-tailed dynamics was strongly correlated with variability in adult survival in time series for Adélie abundance simulated using our age-structured model (Figure C.23). Importantly, Gaussian-distributed adult survival led to non-Gaussian dynamics for abundance. In other words, heavy-tailed distributions for demographic rates are not necessary for heavy-tailed dynamics in abundance. These results provide evidence that the heavy-tailed dynamics identified in regions like the NW AP may be driven by high interannual variability in adult survival rates. Previous studies, including a long-term mark-recapture study on the NW AP [97] and others [70, 193], have found that adult survival rates in the region are characterized by

high variability, most of which could not be explained by environmental conditions but, as suggested by our findings, could be related to variability in large zooplankton (mesozooplankton and macrozooplankton) biomass and NPP (Figure 4.6).

4.5 Conclusions

Irrespective of how populations are counted, Anderson et al. [11] found evidence of heavy-tailed abundance dynamics across 38% of taxonomic orders, including in many insect species and other systems with drastically different life history strategies from the colonial seabird studied here. Thus, the possibility of heavy-tailed dynamics should be considered when simulating time series for any species, especially those subject to extreme events. Conversely, population forecasts incorrectly assuming a Gaussian distribution (when a heavier-tailed distribution would be more appropriate) will be overly optimistic and tend to underestimate extinction risk. We are not aware of any study that quantifies the conservation cost of using a Gaussian distribution to model a system that shows signs of extreme population events, but we believe this is an important research direction. In the meantime, we urge modelers to consider heavy-tailed distributions like the Student's t when making forecasts of population abundance. The Student's t distribution is particularly convenient since its use does not presuppose heavy-tailed dynamics and, in the limit $\nu \to \infty$, approaches the Gaussian exactly. While we focused here on the evidence for, and drivers of, heavy-tailed abundance dynamics, more work is needed to quantify the potential increase in model accuracy with the use of a more flexible distribution for abundance like the Student's t.

Chapter 5

How high interannual variability affects the Living Planet Index: Implications for null models

5.1 Introduction

Developed by the World Wildlife Fund (WWF) and the Zoological Society of London, the Living Planet Index (LPI) [136] measures the overall global trend in vertebrate abundances since 1970. The LPI aggregates population time series for vertebrate species from terrestrial, freshwater, and marine systems into a relative index in which the baseline (1.0) is adjusted to the population measure in 1970. Every two years, the LPI is updated and summarized in the Living Planet Report (LPR) [8]. The LPI has become a prominent indicator of global biodiversity change due to its value to policymakers and its use in assessing the progress toward many biodiversity targets, including the 2011-2020 Aichi Biodiversity Targets set by the Convention on Biological Diversity [50] and the Kunming-Montreal Global Biodiversity Framework, agreed to in 2022 [51]. As the LPI simplifies global species trends across systems to a single, easily-understood metric, it is a powerful tool for practitioners, policymakers, and the public alike.

5.1.1 Calculation of the LPI

The LPI is calculated from the animal population time series data compiled in the Living Planet Database (LPD) [139]. As of 2022, the LPD contains 30,000 population trends for more than 5,200 species of vertebrates (including fish, amphibians, reptiles, birds and mammals) [8]. Since we consider here caveats of the LPI brought on by methodological choices in its calculation, we first describe the procedure for generating the index.

The LPI is calculated using the geometric mean of trends for each species. First, data gaps in empirical population time series are interpolated using either a Generalized Additive Model (GAM) [47] or what is known as the chain method [136]. For population time series with at least six observed counts (with N_t being the observed count in year t), a GAM is fit on the values of $\log_{10}(N_t)$, with the smoothing parameter set to half the number of observed counts for that population [47]. The fitted GAM values are then used to calculate predicted values of N_t (N'_t) for all years from the first to the last observation (including those with observed counts in addition to those with missing data). For population time series with less than six data points, or for which the GAM results in a poor fit, the chain method (log-linear interpolation) is used for interpolation instead [136]:

$$N'_t = N_p(N_s/N_p)^{[(t-p)/(s-p)]}, \tag{5.1}$$

where p is the year preceding t with a measured value and s is the subsequent year with a measured value. In this way, N'_t is calculated for all years between the first and last observed population counts. (For time series in which $N_t = 0$ for some year t, one percent of the mean population for the whole time series is added to all years before interpolation by either the GAM or chain method.)

With fully interpolated time series for each population, annual trends for each species are then calculated as the logarithm of the ratio of population measure for

successive years:

$$d_t = \log_{10}(N'_t/N'_{t-1}), \tag{5.2}$$

calculated for each population time series for all t after the first observed population count, through the last observed count. For species with more than one population time series, the modeled annual trends d_t for each population are then averaged together to obtain a single set of annual trends:

$$\bar{d_t} = \frac{1}{n_t} \sum_{i=1}^{n_t} d_{it}, \tag{5.3}$$

where n_t is the number of populations for that species and d_{it} is calculated as above for population i. Next the index value I is calculated in year t as

$$I_t = I_{t-1} 10^{\bar{d_t}}, \tag{5.4}$$

with $I_0 = 1$ for the initial year, usually set as 1970.

A bootstrap resampling technique is used to generate confidence intervals for the LPI. Each bootstrap replicate is calculated by sampling n_t species-specific values of d_t at random with replacement from the n_t observed values. This is performed for each interannual interval and $\bar{d_t}$ and I_t are calculated as described above. The bootstrapping procedure is implemented 10,000 times, with the bounds of the central 9,500 I values for each year representing the 95% confidence interval for the index in that year [47]. The above procedure to calculate the LPI and associated confidence intervals can be automated using the R package `rlpi` [159].

5.1.2 Known caveats of the LPI

Though the LPI is one of the most popular biodiversity indicators in use, its developers made deliberate trade-offs in the index's design that require careful consideration. First, the LPI standardizes trends in time series as relative rates of change, not absolute population changes. Thus, declines in small populations are viewed as equivalent to similarly proportioned, larger absolute declines in large populations [47]. As a result, the LPI should be interpreted as the average trend in population change, not the average loss in the absolute number of animals [8, 236]. Second, the mathematical characteristics of exponential population growth demand that the LPI be summarized using the geometric mean, which is always lower than the arithmetic mean. Thus the LPI can be disproportionately influenced by even a small number of rapidly declining populations [129].

There are other known weaknesses of the LPI, many of which stem from the data quality of the underlying LPD. In addition to the fact that the LPI does not consider invertebrates or plants, the LPD is also biased towards well-studied species groups and regions. For example, the LPI over-represents temperate regions and under-represents tropical regions [159]. In addition, perhaps predictably, polar regions are vastly underrepresented in the LPI. In fact, the 2022 LPD contains so few data from the Antarctic and Arctic that these regions are entirely omitted from most maps in the 2022 LPR [8]. Nonetheless, new data are frequently added to the LPD and these data gaps continue to shrink with each report [126].

5.1.3 Null models for the LPI

The fact that historical data are being added to the LPD regularly leads to another well-established inconvenience of the LPI framework: indices reported in LPRs from different years cannot be compared, since the underlying datasets differ. In describing this point, the technical supplement of the 2022 LPR notes that the index value in any

given year should always be compared to the initial year of the trendline, the year 1970, in which the index is set to 1 [236]. For this reason, the LPR assumes that the LPI for stable populations will remain at 1.

In a recent study, Buschke et al. [32] demonstrated that symmetric population fluctuations (i.e., caused by demographic or environmental stochasticity, observation error, or ecological drift), even in populations that are stable on average, can have unexpected impact on the LPI. Specifically, Buschke et al. [32] found that small additive population fluctuations led to a declining LPI despite average population sizes remaining steady since the logarithmic transformation used to calculate the LPI generates an asymmetrical response to equal positive and negative fluctuations. The authors argue that, although the observed declines in the LPI attributed to population fluctuations are small compared to the 69% global decrease reported in the 2022 LPR, the "null expectation of the LPI should be a declining counterfactual rather than a static baseline set at 1970" [32]. This study's findings do not invalidate the LPI as an effective tool for measuring biodiversity, but instead encourage the use of randomized null models when considering the significance of LPI trends.

In order to illuminate simply a heretofore unrecognized issue, Buschke et al.'s analysis [32] only considered the effect of small additive population fluctuations to otherwise stable populations. However, some animal species display interannual fluctuations that are markedly larger than the variation considered by Buschke et al. [11, 39, 59, 129], including Antarctic penguins of the genus *Pygoscelis* [218]. Here we continue and expand on the work started by Buschke et al. [32] to investigate these various issues, which impact the LPI most acutely when interannual fluctuations are large, and highlight the difficulties in simulating time series with large additive interannual fluctuations that are truly stable on average. We then argue that population fluctuations are better characterized by treating abundance on a logarithmic scale rather than on a linear scale (as in Buschke et al. [32]), and discuss the resulting null model for the LPI.

5.2 Buschke et al.'s approach: Constant additive population fluctuations

5.2.1 Original simulations

Buschke et al. [32] simulated how random fluctuations in population time series affect the empirical global LPI by iterating three null models with positive or negative interannual fluctuations (1, 3, and 5% fluctuations annually). These simulations included the same number of time series as the 2020 empirical LPI [9], with starting population counts identical to those in the associated database. Each population time series was simulated with additive interannual fluctuations equal to a constant percentage of the original population count:

$$N_t = N_{t-1} \pm (N_0 * a), \tag{5.5}$$

where N_0 is the initial observed count for the population and $a = [1, 3, 5\%]$ is the level of population fluctuation as a percentage of the initial population count. Thus the number of individuals (i.e., the quantity being counted, which may be individuals, breeding pairs, nests, etc.) that the population fluctuates by in each year is constant. For each time series, the annual direction of the fluctuation was selected randomly with positive and negative fluctuations being equiprobable. Simulated populations were allowed to fluctuate annually between the first observed measurement in the time series to the end of the time series, and then thinned to match the data gaps in the empirical time series. Importantly, if a simulated population time series had years in which the generated value of N_t was negative, the entire time series was replaced with missing data, removing the population entirely from consideration for the calculation of the LPI for that iteration. The LPI was then calculated as in [159] for each simulation, with the entire process iterated 100 times for each of the three levels of fluctuation (1,

3, 5%).

Here, we aim to follow the above methodology of Buschke et al. [32] for a different underlying dataset, consisting of 328 time series of all known Chinstrap penguin (*Pygoscelis antarcticus*) breeding populations and spanning from 1970-2019, compiled and organized by the Mapping Application for Penguin Populations and Projected Dynamics (MAPPPD; [102]). This dataset is used for illustrative purposes and computational ease rather than the much larger LPD used by the authors of the original study [32]. We use these Chinstrap time series to explore the performance of Buschke et al.'s LPI null model framework for populations displaying larger interannual fluctuations. Like other penguin species of the genus *Pygoscelis*, Chinstrap penguins are known to exhibit high interannual variability in observed population time series (see Appendix E.1.1; [218]). To allow for the higher interannual fluctuations as seen in some animal species [39, 59, 129], we consider more levels of fluctuation allowing $a = [1, 3, 5, 7, ..., 31, 33\%]$.

5.2.2 Performance of Buschke et al.'s simulation framework for larger population fluctuations

Before calculating the LPI for simulated Chinstrap penguin time series, we first explore the performance of the above simulation framework (Equation 5.5) when population fluctuations take on a larger range, $a = [1, 3, 5, 7, ..., 31, 33\%]$, for a toy example of 500 simulated time series of equal initial population count (following Buschke et al., Figure 1 [32]). We set the initial population count to 1000 to represent a medium-sized Chinstrap penguin colony (the median initial observation in Chinstrap penguin breeding population time series is 1200 nests).

For populations with large additive interannual fluctuations following the above framework, simulated population counts can easily become negative when population fluctuations are more than 5% of the initial population count (Figure 5.1b). Following

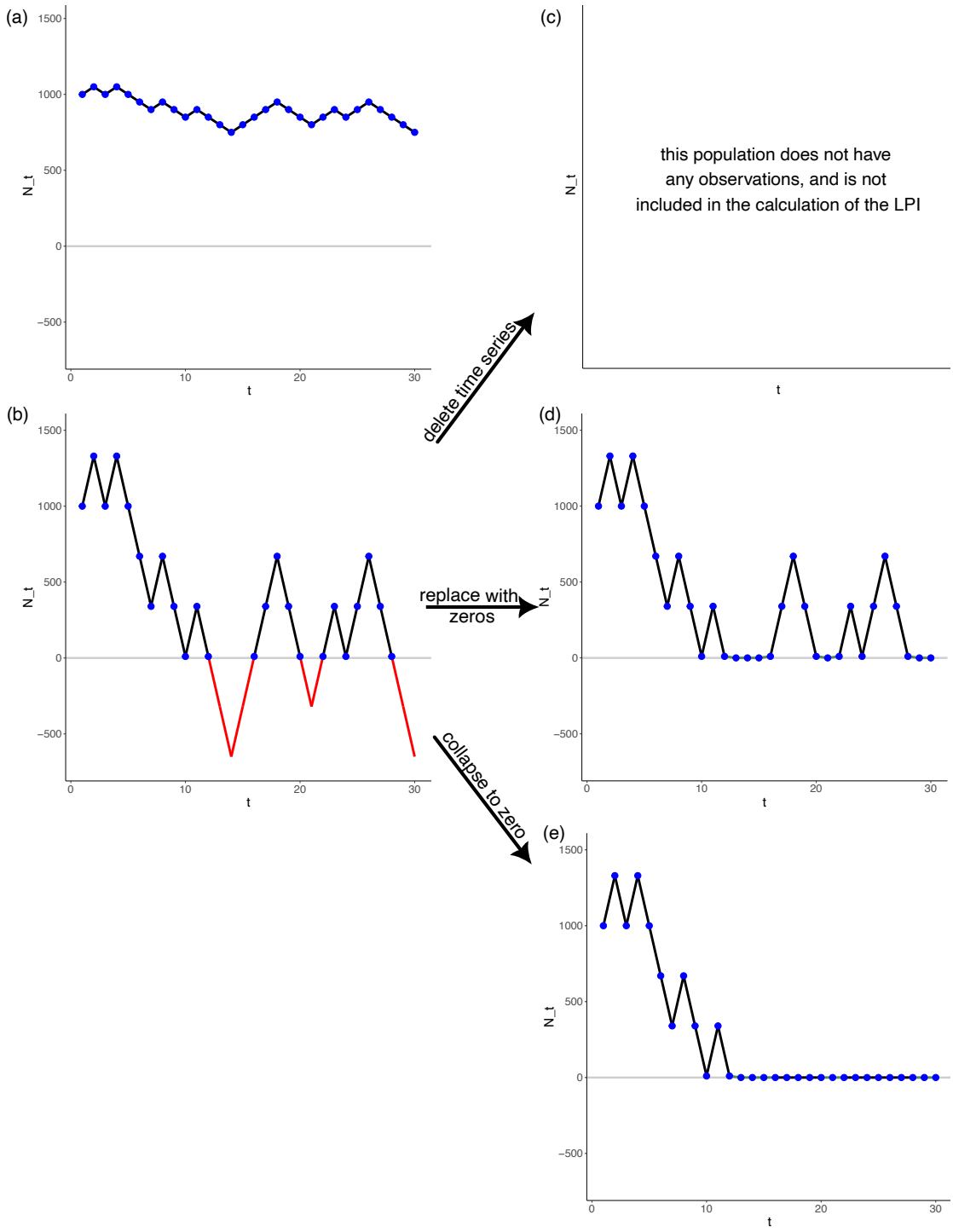

FIGURE 5.1: Simulated population time series with constant population fluctuations as a (a) small (5%) or (b) large (33%) percentage of the initial population count. When fluctuations are high (b), simulated population counts can easily become negative. Buschke et al.'s original framework [32] replaces time series with negative population counts entirely with missing data (c), deleting the time series from consideration in any calculation of the LPI. Other options for eliminating negative population counts include (d) replacing negative counts with zeros and (e) collapsing the time series to zero for all remaining years beginning with the first negative count.

Buschke et al.'s framework [32], time series with negative population counts are entirely replaced with missing data, removing those populations from consideration for the calculation of the LPI (Figure 5.1c). Thus these decreasing populations are deleted from the collection of time series, but populations that are symmetrically increasing in the same way remain, leaving populations that are increasing on average instead of being stable as desired. Figure 5.2 shows the average percent change in population count from the initial population of 1000 to the final population count for 500 simulated time series. On average, populations grow by more than 30% under this simulation framework when population fluctuations are 15% of the initial population count (simulation framework 1, Figure 5.2). When fluctuations are equal to 30% of the initial count, populations grow by over 100%. To accurately represent the state of populations in the absence of true growth or decline in a null model, populations must be stable on average. Thus, when population fluctuations are believed to be additive and above 5%, we must find an alternative simulation framework for use in a null model for the LPI.

Since this problem (i.e., simulated populations not being stable on average as desired for a null model) stems from simulated time series that drop below zero being deleted from consideration, we consider other procedures for dealing with simulated time series with negative population counts. If simulated time series were permitted to have negative values of N_t, populations would be completely stable on average, since largely increasing populations would be equally balanced by largely decreasing populations. Since population counts must be nonnegative, however, we instead consider the case where negative simulated values are replaced with a value of zero. In this case (Figure 5.1d), time series are not removed from consideration when they include negative values, but instead they are remedied by replacing those negative values with zeros. While this framework does reduce the problem of bias in simulated population time series, populations still grow on average (by up to 27%) when population fluctuations are more than 15% on average (simulation framework 2, Figure

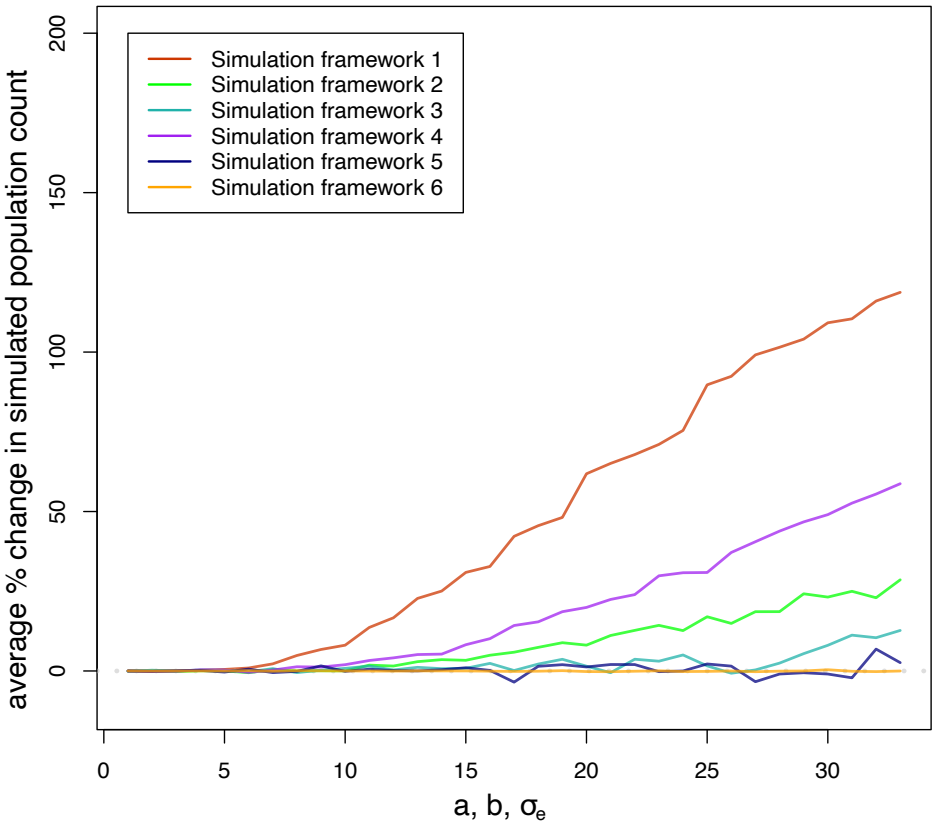

FIGURE 5.2: Average percent change in population count from the initial population count after 52 years for varying levels of population fluctuations (a, b, or σ_e depending on the simulation framework, see Section 5.4.1). Simulation framework 1: Equation 5.5 with time series simulated to have negative population counts removed (see Figure 5.1c) [32]; Simulation framework 2: Equation 5.5 with simulated negative population counts replaced with zeros (see Figure 5.1d); Simulation framework 3: Equation 5.5 with time series collapsing to zero when simulated population counts are negative (see Figure 5.1e); Simulation framework 4: Equation 5.6; Simulation framework 5: Equation 5.7; Simulation framework 6: Equation 5.8.

5.2).

Instead of replacing negative simulated population counts with zeros, we can instead collapse time series to zero and make them unable to recover when population counts reach negative values (Figure 5.1e). Under this framework, populations are much closer to stable on average, with the average population only growing steadily

when additive population fluctuations are more than 28% of the initial count (simulation framework 3, Figure 5.2). Thus none of these simulation schemes can generate populations that are stable on average when additive population fluctuations are believed to be over 28% of the initial population count.

Note that treating individual negative abundance values in simulated time series (as in Figure 5.1b) as missing values results in similar challenges when attempting to generate stable populations on average in addition to creating multiple other confounding issues when attempting to calculate the LPI (see Appendix D.1).

5.3 Alternative approaches: Random additive population fluctuations with a zero-truncated normal distribution

A straightforward way around the decision between the options in Figure 5.1 might be to draw population fluctuations from a truncated distribution to avoid simulating negative population counts in the first place. Population fluctuations drawn from a zero-truncated normal distribution, instead of being held constant at a percentage of the initial population count, would never result in simulated time series with negative population counts. We repeat the above methodology in a separate suite of simulations, with annual population fluctuations for each time series now drawn from a truncated normal distribution,

$$N_t \sim \text{trunc}N(\mu = N_{t-1},\ \sigma = N_0 * a,\ 0, \infty), \tag{5.6}$$

where $a = [1, 3, ..., 33\%]$ is still the level of population fluctuation as a percentage of the initial population count. (Note the use of $N_0 * a$ as the standard deviation of this distribution; in a non-truncated normal distribution with a standard deviation of $N_0 * a$,

the probability that the interannual fluctuation in a given year is less than or equal to $N_0 * a$ is 68%, with 95% of fluctuations less than or equal to $N_0 * 2a$ [18].) Since the zero-truncated normal distribution will not produce negative simulated population counts, this simulation framework does not require replacing negative population counts in any way. However, since fluctuations drawn from a truncated normal distribution are not symmetric, time series simulated with a zero-truncated normal distribution grow significantly on average for population fluctuations when $a > 5\%$ (simulation framework 4, Figure 5.2), with problems almost as severe as Buschke et al.'s original framework that deletes time series that fall below zero at any point (simulation framework 1, Figure 5.2). Thus this simulation framework does not provide a method for simulating stable populations when additive population fluctuations are high.

5.3.1 Shifting fluctuation benchmarks

We next consider random population fluctuations that are centered on a percentage of the previous year's population count instead of the initial population count, repeating the above simulations with annual population fluctuations for each time series now drawn from the following truncated normal distribution:

$$N_t \sim \text{trunc}N(\mu = N_{t-1},\, \sigma = N_{t-1} * b,\; 0, \infty), \tag{5.7}$$

where $b = [1, 3, ..., 33\%]$ is the level of population fluctuation as a percentage of the population count in the previous year (not the initial population count). Note that using the zero-truncated distribution allows us to use this shifting benchmark here since N_{t-1} is guaranteed to be a positive number—a similar shifting benchmark would not be possible under a simulation framework defined by Equation (5.5).

When interannual fluctuations are drawn from a zero-truncated normal distribution centered on a percentage of the previous year's population count, simulated populations are roughly stable on average for population fluctuations up to 33% of the

previous count (simulation framework 5, Figure 5.2). Under this framework, simulated time series that have grown to a point much larger than the initial population count are more likely to increase or decrease by a large amount in the following year than they would be if population fluctuations were always benchmarked to the original count (as in Equation 5.6). On the other hand, simulated time series that have fallen close to zero will fluctuate much less on average. Thus time series that undergo a marked increase in population count have a considerable chance of reverting back to a level near the initial count, and time series that undergo a marked decrease have a harder chance recovering to larger counts. On average, simulated time series under this framework are therefore much closer to stable than those with population fluctuations that remain benchmarked to the initial population count. This marks the most successful framework for simulating stable on average population time series with additive population fluctuations that we have discussed here.

With this simulation framework that yielding stable population time series on average, we next calculate the LPI for simulated Chinstrap penguin breeding colonies to quantify the null expectation of the LPI for populations undergoing these additive fluctuations. For 100 iterations, we simulate each of the 328 Chinstrap breeding abundance time series under the above framework (drawing population counts from a zero-truncated normal distribution with fluctuations centered on the previous year's count, as in Equation 5.7) and calculate the LPI as described above. The resulting LPI in each case decreases on average over the 100 iterations, with a 96% decrease in the LPI after 52 years when fluctuations are 33% of the value of the previous count (Figure 5.3). For small fluctuations (1-5%), similar levels of decline in the LPI were found to those described in the original study [32] (Figure D.1). With the LPI's design meant to accommodate multiplicative population dynamics, positive additive population fluctuations (increases) cannot offset negative population fluctuations (declines) of equal size. The extent of the resulting decline in the LPI is proportional to the level of population fluctuations experienced by the population time series (Figure 5.3), with

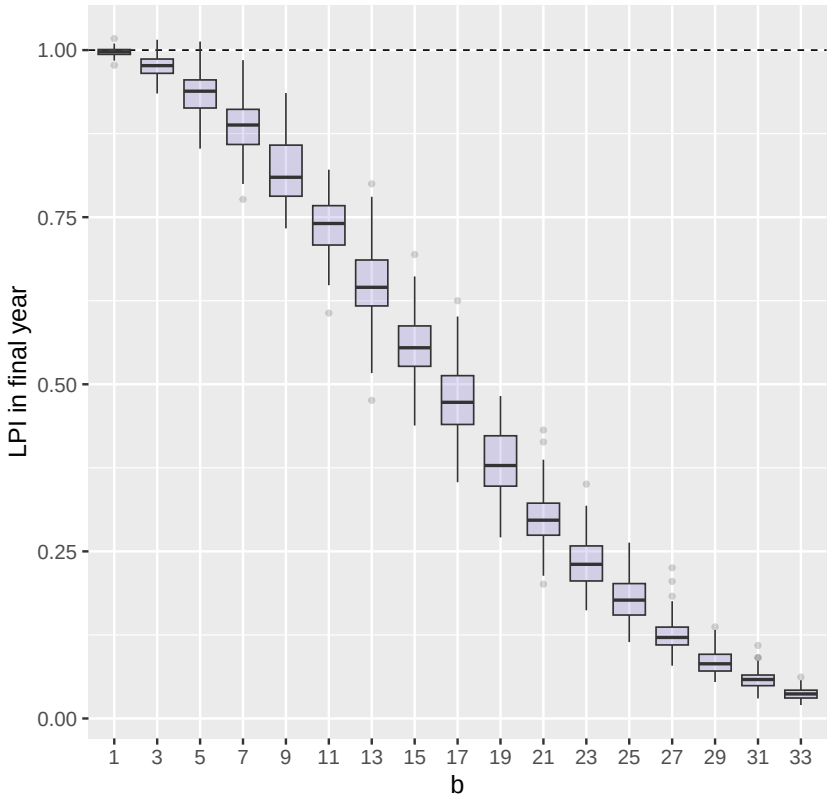

FIGURE 5.3: Final-year LPI for Chinstrap penguin populations simulated to be stable on average with drawn from a truncated normal distribution with varying levels of population fluctuations ($b = [1, 3, ..., 33\%]$). Random population fluctuations are benchmarked to the population count in the previous year (see Equation 5.6).

population fluctuations of more than 20% resulting in a null expectation of the LPI that is actually lower than the 2022 empirical LPI [8]. Thus, when population time series are believed to undergo large additive fluctuations in the absence of true growth or decline, the null expectation of the LPI is a declining counterfactual—and one that is potentially quite steep. This is a continuation of the trend seen in Buschke et al. [32] for populations experiencing low additive population fluctuations.

5.4 Alternative approaches: Random logarithmic population fluctuations

As we have shown, simulating time series with large population fluctuations that are stable on average is difficult when fluctuations are assumed to be additive. When population fluctuations are this large, however, interannual variability is often best understood to exist on the logarithmic scale instead of the linear scale [119, 168, 171, 225]. We now repeat the above simulations with non-additive random population fluctuations drawn from a log-normal distribution and present a null model framework for the LPI with these logarithmic population fluctuations. Specifically, we simulate population time series fluctuating on the logarithmic ($\log_e$) scale as:

$$\log(N_t) \sim N(\log(N_{t-1}),\ N_0 * \sigma_e), \tag{5.8}$$

where σ_e is the standard deviation on the $\log_e$ scale. Bayesian state-space models for Chinstrap penguin population dynamics [217] have estimated a process error of $\sigma_e = 0.33$ (see Appendix E.1.1). As such, we use the range $\sigma_e = 0.01, 0.03, ..., 0.33$ here. Since simulating time series on the logarithmic scale ensures that populations are never negative, we do not have to replace negative values with either missing data or zeros.

Population time series simulated on the logarithmic scale are stable on average regardless of how large the level of population fluctuation σ_e (simulation framework 6, Figure 5.2), with the issues encountered on the linear scale resolved. Thus, when stable population time series with large interannual fluctuations are required, as for use in a null model for the LPI, logarithmic population dynamics make their generation straightforward.

Therefore we next use this simulation framework to calculating the LPI for simulated Chinstrap penguin breeding populations, now with each population time series

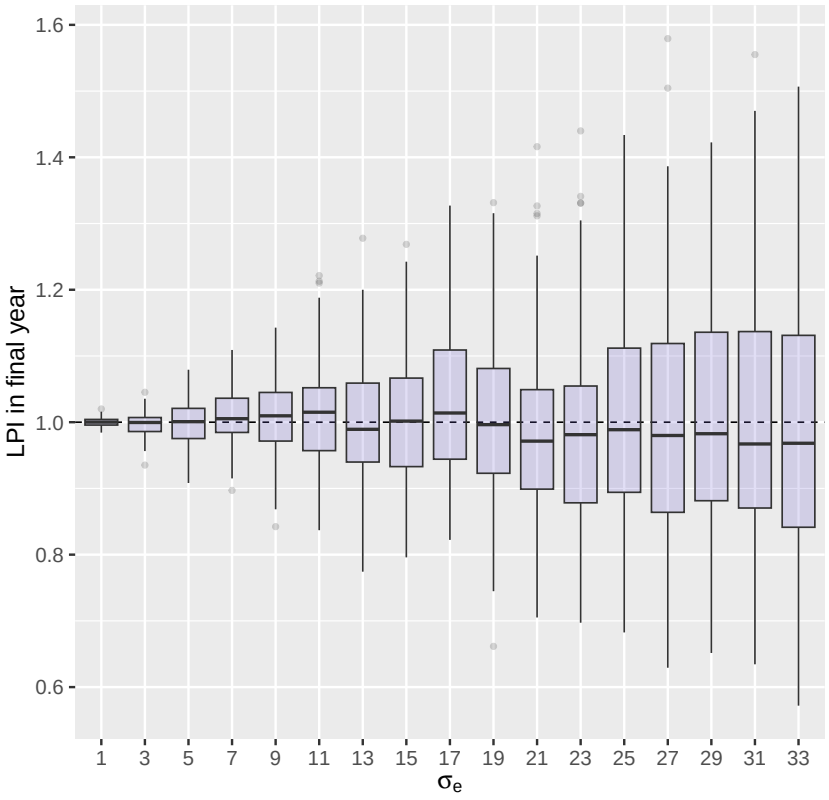

FIGURE 5.4: Final-year LPI for Chinstrap penguin populations simulated to be stable on average with drawn from a log-normal distribution with varying levels of population fluctuations ($\sigma_e = 0.01, 0.03, ...0.33$, see Equation 5.8).

following a log-normal distribution as described in Equation (5.8). As above, we calculate the LPI for 100 iterations for each level of population fluctuations. Even when time series undergo very large population fluctuations on the logarithmic scale, the value of the LPI after 52 years is close to 1 on average across 100 iterations (Figure 5.4). Thus, when populations are assumed to vary on the logarithmic scale instead of the linear scale the null expectation of the LPI is not declining, but instead a virtually stable line at 1. Since the LPI is designed for use with multiplicative population dynamics, fluctuations simulated on the logarithmic scale are symmetric in LPI calculations.

5.4.1 Translating logarithmic fluctuations to the linear scale

It is worth investigating how population fluctuations on the logarithmic scale compare to those simulated on the linear scale. For a given σ_e, we can calculate on average how much a population with multiplicative fluctuations is varying annually on the linear scale (i.e., the change in raw count). On the $\log_e$ scale, an increase of $1\sigma_e$ translates to a $(e^{\sigma_e}-1)\%$ increase in raw count and a decrease of $1\sigma_e$ translates to a $(1-e^{-\sigma_e})\%$ decrease in raw count. Since these changes are not symmetric on the linear scale, we take the geometric mean of these two quantities to estimate the percent change in raw count for a change of $1\sigma_e$ on the logarithmic scale. Using this translation, $\sigma_e = 0.33$ is equal to a 33.1% change in raw count and a $\sigma_e = 0.01$ translates to an approximately 1% change in raw count. Thus a standard deviation of σ_e for the log-normal distribution in Equation (5.8) represents approximately a $100\sigma_e\%$ change in population on the linear scale. Since Chinstrap penguin populations are estimated to have a process error of $\sigma_e = 0.33$, we have used the ranges $a,\, b = [1, 3, ..., 33\%]$ and $\sigma_e = [0.01, ..., 0.033]$ to represent comparable ranges in population fluctuation levels across simulation frameworks.

5.5 Discussion

The LPI is made up of thousands of populations that each have different levels of underlying population fluctuations due to demographic or environmental stochasticity, ecological drift, or observation error. Understanding how the LPI is affected by these random fluctuations when populations are stable on average is crucial to the interpretation of trends in the index. As the framework of the LPI is gaining momentum and is increasingly being used to quantify population changes in specific taxa [174, 192], habitats [160], and countries [57, 209], having an accurate and robust null expectation for index values is critical.

To produce an accurate null model for the LPI, simulated population time series

should be produced under a framework that closely describes the believed underlying dynamics when populations are neither growing nor declining. Fundamentally, the time series simulated should be stable on average. As we have shown, many potential frameworks for simulating populations with additive fluctuations will not produce stable populations on average when fluctuations are large. Constant additive population fluctuations greater than 5% of the initial population growth will result in simulated time series that are growing on average since population counts cannot be negative. Though there are many choices for how to account for negative simulated values (Figure 5.1), none of them form a simulation framework that yields stable populations when population fluctuations are additive, constant, and more than 25% (Figure 5.2). When simulated time series are stable on average, they are suitable for use in the calculation of a null model for the LPI, but we have shown that this is not always the case. When population fluctuations are not constant but instead drawn randomly from a truncated normal distribution, stable populations can only be generated for interannual fluctuations of more than 5% when those fluctuations are benchmarked to the previous year's population instead of the initial count. Only when population fluctuations shift their baseline to the previous year's population count can large additive fluctuations can be large with simulations still accurately generating stable populations on average.

Even when populations display sufficiently low interannual fluctuations for a simulation framework to produce stable populations, it is important to carefully consider if this null model is accurately representing what practitioners believe to be the null expectation of population dynamics in the absence of meaningful growth or decline. We appreciate the light Busckhe et al. brought to this issue with their work [32], as understanding how the LPI is affected by random population fluctuations is crucial to the interpretation of trends in the index. We disagree, however, with those authors' focus on additive population fluctuations, which may represent a distracting strawman for evaluating the performance of the LPI. Buschke et al. [32] argue that ecological

drift, demographic stochasticity, and observation error are additive processes. Under the assumption of a closed population, however, demographic processes do not result in a constant number of births or deaths irrespective of the current population count. Overwhelmingly, ecologists instead model growth, death, and survival as multiplicative processes, resulting in abundances that fluctuate log-normally [119, 168, 171, 225]. For these reasons, the LPI was designed to consider populations undergoing exponential dynamics. When population fluctuations are assumed to be multiplicative, the resulting null model for the LPI shows no bias (the null expectation of the LPI is, in effect, a constant line at 1; see Figure 5.4).

Additive fluctuations in observed animal populations would only reasonably model immigration and emigration, and would thus be very small relative to the population count. A population with large additive fluctuations would hardly be a closed population, and careful consideration would need to be taken if one hoped to include them as a population in the LPI framework. While populations may experience a combination of additive and multiplicative fluctuations, any additive population fluctuations would be outweighed by the fluctuations in the multiplicative birth, death, or survival rates. As we have shown, even large multiplicative fluctuations do not bias the LPI in a significant manner, leaving a null expectation for the LPI that is virtually stable at 1 throughout the time series; thus, the LPI is not inherently biased, as may be concluded by the casual reader of Buschke et al.'s original paper [32].

5.5.1 Consequences of the LPI's GAM and Buschke et al.'s suggested reshuffling null model

Buschke et al. [32] also show in their original analysis that the use of a generalized additive model (GAM) to interpolate and smooth empirical time series leads to biased LPI estimates when populations are increasing or decreasing non-linearly. In light of this, the authors suggest another randomized null model in which the starting and

ending population counts in time series are maintained but the order of the incremental changes to the population seen in the empirical time series are randomized. Each resulting final-year LPI value would represent a potential trajectory of the population to the same measured end-state, averaging out the bias that Buschke et al. [32] identified when the GAM is fit to nonlinear trajectories.

We completely agree with Buschke et al. [32] that the use of the GAM to interpolate time series in the context of the LPI is problematic. In fact, given the title of their original paper, readers of Buschke et al.'s study may not fully appreciate that the issue lies with the use of the GAM to interpolate missing data, not with the mathematical framework of the LPI. The GAM has two additional down-stream impacts on the LPI that were not emphasized in the original study. For one, the GAM is highly sensitive to the density of the empirical data and even minor changes in the data (particularly at the beginning and ends of the time series) can have drastic impacts on the final estimated GAM and the resulting LPI. Secondly, the GAM not only interpolates missing data, but also smooths the time series where data were observed. This smoothing of the data leads to a drastic reduction in interannual variation that has a substantive impact on the resulting index. Since the abundance time series of some species, like Antarctic penguins, display considerable interannual fluctuations [11, 39, 218], preserving this variability is critical to understanding their population dynamics and producing an accurate index of change through time.

We disagree with Buschke et al.'s [32] suggested approach of using a randomized reshuffling null model to combat the bias introduced by the LPI. Not only does this reshuffling null model not address the additional problems with the GAM that we noted above, but—as noted by Buschke et al. themselves [32]—this suggested approach cannot be used to estimate the trend of the LPI across the whole time series, and instead requires a narrowing of focus to only the value of the LPI in the final year of the time series (at the expense of how changes in populations might accrue

through time). Calculating the full trajectory of the LPI can yield clear epochs of population change that may link to ecological or anthropogenic drivers [47, 136, 174], but these nuances are lost when reshuffling fluctuations between the beginning and ending abundances. We contend that a better approach would be to simply replace the GAM with a more mechanistic framework for data interpolation, and suggest a state-space model (SSM). The use of an SSM [39, 114, 174] addresses all of the above concerns with the GAM without requiring any ad hoc post-processing and preserves the capacity to look at change within the time series rather than simply at the two end points. SSMs also allow for a much more informed modeling approach to improve interpolation of missing data than that provided by a GAM, and easily facilitate the incorporation of covariates, spatial autocorrelation, or even age-structure. As the LPI grows in popularity as a biodiversity index for use by policymakers, addressing these key statistical issues becomes of increasing importance.

It is important that careful consideration is taken when deciding on a null model for the LPI. If true population fluctuations are unknown or difficult to estimate—or if the LPI is being calculated for diverse species and resources do not allow for thoughtful attention to be paid to each one individually—several alternative null models may be necessary for comparison to interpret true decline or growth. As Buschke et al.'s work shows and our efforts expand on, the worst choice for a null model for the LPI is would be failing to develop a null model entirely, since an accurate understanding of the index first requires a thoughtful examination of the nature of the underlying time series and their intrinsic stochasticity.

Chapter 6

Penguindex: A biodiversity indicator for *Pygoscelis* species penguins identifies key eras of population change

6.1 Introduction

As summarized in the most recent Global Biodiversity Outlook from the Convention on Biological Diversity, none of the 20 Aichi Biodiversity targets set for 2011-2020 were fully met [50], and the Antarctic, often considered to be buffered from anthropogenic disturbance, is not fairing any better [42]. The Kunming-Montreal Global Biodiversity Framework agreed to in 2022 includes a new set of goals and targets for which progress must be measured [51], so quantifying the changes in global biodiversity remains one of the most important ecological endeavors today. Understanding ecological change is especially urgent for systems in which changes are occurring more rapidly. Among these is the Antarctic and its Southern Ocean ecosystem, which is experiencing significant warming and the resulting changes in sea ice distribution, shifting winds, and increased ocean acidification [123, 175, 188, 228, 227, 241]. These changes have influenced the intricately connected Southern Ocean food webs in countless ways, affecting the success, abundance, and distribution of many species [5, 72,

234]. It is particularly difficult, however, to assess ecological change in the Antarctic; complications include separating natural variability from shifting regional trends, inadequate historical and current data on both terrestrial and marine diversity, and logistical challenges to science in remote regions.

As important marine predators in the Southern Ocean ecosystem, penguins of the genus *Pygoscelis* (Adélie penguins, *P. adeliae*; Chinstrap penguins, *P. antarctica*; and Gentoo penguins, *P. papua*) are critical bellwethers of climate change and, as a result, serve as an ideal focus for investigations into ecological change in the Antarctic. Over the last decade, there has been a concerted effort to catalog the distribution and abundance of each of the three *Pygoscelis* species penguins in the Antarctic (south of 60°S), including several efforts to use satellite imagery to complete pan-Antarctic population censuses for each species [95, 141, 210]. In addition, the completion of the Mapping Application for Penguin Populations and Projected Dynamics (MAPPPD; [102]) now provides easy access to all publicly available census data dating back to 1979 [38]. MAPPPD's release has facilitated a renewed interest in continental scale penguin dynamics that has uncovered differing trends across pygoscelids in response to climate change [39, 198]. However, until now, population trends of pygoscelid penguins have not been synthesized into a single global indicator for use by policymakers. The need for easy-to-interpret metrics of penguin trends has never been more critical, as the Antarctic Treaty Parties address threats of the changing climate and increased human activities, and the Convention on the Conservation of Antarctic Marine Living Resources (CCAMLR) wrestles with the design (and eventual evaluation) of Marine Protected Areas [22, 164]. In an effort to meet this urgent need, we introduce here a pygoscelid penguin-specific biodiversity index, the 'Penguindex,' using the framework of the Living Planet Index (LPI).

The LPI, a global biodiversity index produced by the World Wildlife Fund and the Zoological Society of London, is a major collaborative effort to track trends in vertebrate abundance around the globe [8]. The index aggregates individual time series

of vertebrate population measures to track average changes in abundance of species over time [47, 136, 159]. The biennial Living Planet Report (LPR) uses the LPI to distill global biodiversity trends into a singular message on the health of our planet. The 2022 LPR describes a 69% average decrease in global biodiversity since 1970 [8]. The LPI has also been used to identify trends across many taxa, and several country- [150, 209], biome- [75, 160], and taxa-specific [93, 192, 174] sub-indices have been developed to allow for easy-to-understand monitoring of biodiversity. Our Penguindex leverages this methodology for pygoscelid penguins, and fills a critical gap in the monitoring of biodiversity change in the Antarctic.

While penguin population data have been collected and analyzed for decades, the Antarctic community has not made a concerted effort to integrate those data into global biodiversity efforts such as the LPI. As a result, the Antarctic is vastly underrepresented in the database underlying the global LPI [139], and little attention has been drawn to how this might influence global patterns. The 2022 LPR acknowledges that "polar regions ... showed the highest impact probabilities for climate change, driven in particular by impacts on birds" (page 41, [8]), but there is no specific mention of the Antarctic in the report. Though the rate of new Antarctic time series added to the LPI database has accelerated in recent years [126], the data within MAPPPD has not been integrated into the LPI database and the LPI's coverage of Antarctica remains inadequate.

We present the Penguindex as a pygoscelid-specific LPI and an easily-interpreted measure of penguin trends in the Antarctic. Including almost every known pygoscelid breeding site, we first use a Bayesian state-space framework [39] to estimate trends in the relative abundance for all three pygoscelid species, allowing us to leverage experience modeling penguin abundance to more accurately interpolate gaps in observed time series. Using these trends, we calculated the global Penguindex for pygoscelid penguins by aggregating over each of eight Antarctic regions for each species, calculating both species-specific indices and region-specific indices for each species along

the way. In evaluating these trends we also estimate change points—that is, points in time where an index curve shifts significantly—in an effort to understand the mechanisms of these changes. We anticipate that this Penguindex will represent the latest significant development in the monitoring of these important sentinel species of climate change.

6.2 Methods

6.2.1 Population time series

Data on nests and chicks for the three pygoscelid species were collected and organized under the auspices of MAPPPD [102]. We included data from all known breeding populations with at least one observed abundance count between the 1970/1971 season (hereafter referred to as the 1970 season) and the 2019 season, totaling 271 Adélie, 358 Chinstrap, and 109 Gentoo penguin populations with a total of 3884 observed counts. These data were used to fit a Bayesian state-space model to estimate annual pygoscelid nest abundances for each breeding population from 1970 – 2020. This hierarchical model, adapted from [39], included observation error (uncertainty in the number of true nests counted in each year) and process error (stochastic variability in the population growth rate); we modeled the intrinsic rate of growth $r_{i,t}$, for the i_{th} population in the t_{th} season as a function of site and season effects. Notably, the Penguindex does not yet include the Emperor penguin because the data available are currently too patchy.

6.2.2 Calculating the Penguindex

Data subsetting and Antarctic regions

In calculating the Penguindex, since few abundance counts are available prior to 1979 or for the year 2020 (due to COVID-19), we restrict our calculation to the 1980-2019

seasons. While the Bayesian state space model provides estimates for all years for all populations, here we follow the criteria for inclusion of time series in the global LPI [47] and discard from consideration those populations with fewer than two observed abundance counts from 1980 to 2019. This filtering results in 118 Adélie penguin, 94 Chinstrap penguin, and 58 Gentoo penguin populations with fully interpolated time series from which to calculate the Penguindex. Following the LPI framework [47], one percent of the mean population for the whole time series was added to years in time series for which the Bayesian state space model assigned a population value of zero (as was the case for populations with confirmed absence for that species) in any year.

Each breeding population is assigned a geographical region of the Antarctic: (1) Central-west Antarctic Peninsula (AP) and Northwest AP; (2) Southwest AP; (3) Elephant Island, South Orkney Islands, and South Shetland Islands; (4) Northeast AP; (5) Ross Sea (CCAMLR Subareas 88.1 and 88.2); (6) Bellingshausen Sea (CCAMLR Subarea 88.3); (7) Northeast Antarctica (CCAMLR Division 58.4.1); and (8) Southeast Antarctica (CCAMLR Division 58.4.2). The locations of the breeding populations for each species are shown, differentiated by region, in Appendix S1: Figures S1-S3. Adélie penguins are found in all eight regions but Chinstraps breed only in Regions 1-3 and 5, and Gentoos in Regions 1, 3, and 4.

Calculation of index values

The Penguindex is calculated following the general format of the LPI [47, 159] (see Appendix S1: Figure S4). The global pygoscelid index is calculated by aggregating over all three species, each of which is first aggregated over each region. For each breeding population, the annual rate of change d_t is the logarithm of the growth rate in a given year t, $d_t = \log_{10}(N_t/N_{t-1})$, where N_t denotes a draw from the posterior for nest abundance in year t as estimated from the Bayesian state-space model. Drawing from the posterior allows us to propagate the uncertainty regarding abundance in years t and

$t-1$ to the estimate of d_t. For each year t, the values of d_t for each breeding population in a region is then averaged within each species $\times$ region combination, with each breeding population weighted equally, yielding a region-and-species-specific estimate of d_t for each year. These region-specific interannual changes were then aggregated to obtain a single annual rate of change for each species, with each region's interannual change weighted by the proportion of the total number of that species' breeding populations that occur in that region. Annual trends for each species were then aggregated to obtain a single global annual rate of change for all pygoscelid species, $\bar{d}_t$. All three pygoscelid species were weighted equally for this aggregation. Each of these regional, species, and global trends were then converted to index values, $I_t = I_{t-1} \times 10^{\bar{d}_t}$, with $I_{t=1} = 1$ for the reference year 1980. Indices are calculated for each of 1,000 draws from the posterior distributions of nest abundance estimates from the Bayesian state-space model. These analyses were performed using R v4.1.2 [180].

Penguindex null models

Random fluctuations in time series, even when overall population trends are stable, can disproportionately affect the LPI relative to actual trends [31]. Additionally, the population dynamics of *Pygoscelis* penguins are characterized by large interannual fluctuations [39, 240]. To account for this potential bias in the Penguindex, we used a null model that maintained the starting populations in each time series and simulated stable dynamics with random fluctuations. For each species, the posterior mean for the species-specific process error σ was used and the abundance of population i in year t, $N_{sim_{i,t}}$ was simulated as:

$$\log(N_{sim_{i,t}}) \sim \text{normal}(\mu_{i,t} = \log(N_{sim_{i,t-1}}), \sigma^2_{spp}) \tag{6.1}$$

where $N_{sim_{i,t=1}}$ is drawn from the posterior distribution for nest abundance for population i in year $t = 1980$ as estimated from the Bayesian state-space model. For Adélie

penguins, the process error σ depends on the region. These null model time series are then used to calculate the Penguindex as described above. We iterate this null model 1,000 times and average the index over all iterations, obtaining a null index for each region, species, global index as above. This index can then be used as a null expectation of the Penguindex rather than the static baseline of $I_{1980} = 1$.

Era identification

Change points in the Penguindex were identified via segmented regression [166] to find years at which the linear trend of the index changed significantly. This change point analysis allowed us to establish eras of pygoscelid population dynamics between 1980 and 2019. Change points were identified for the global *Pygoscelis* Penguindex as well as for each species- and region-level indices. These analyses were performed using R v4.1.2 [180] and the package `segmented` [167]. The Bayesian information criterion (BIC) was used to select the number change points between 0 and 10 [222]. The maximum number of change points allowed in this procedure was set to 10 since the optimal number of change points never reached the limit of 10 for any index.

6.3 Results

6.3.1 Global *Pygoscelis* trends

On average, *Pygoscelis* penguin populations have grown by 16.9% (95% credible interval, 12.8-21.1%) between 1980 and 2019 (Figure 6.1). Our data suggest an initial surge of growth prior to 1986 (1986 index 1.098, 95% CI = 1.058-1.135), followed by a period of global stability until 1996 (1996 index 1.095, 95% CI = 1.055-1.140); global *Pygoscelis* penguin populations steadily grew between 1996 and 2015 (2015 index 1.206, 95% CI = 1.166-1.249), after which the average population declined suddenly. The null model for the global *Pygoscelis* Penguindex stayed steady at 1.0.

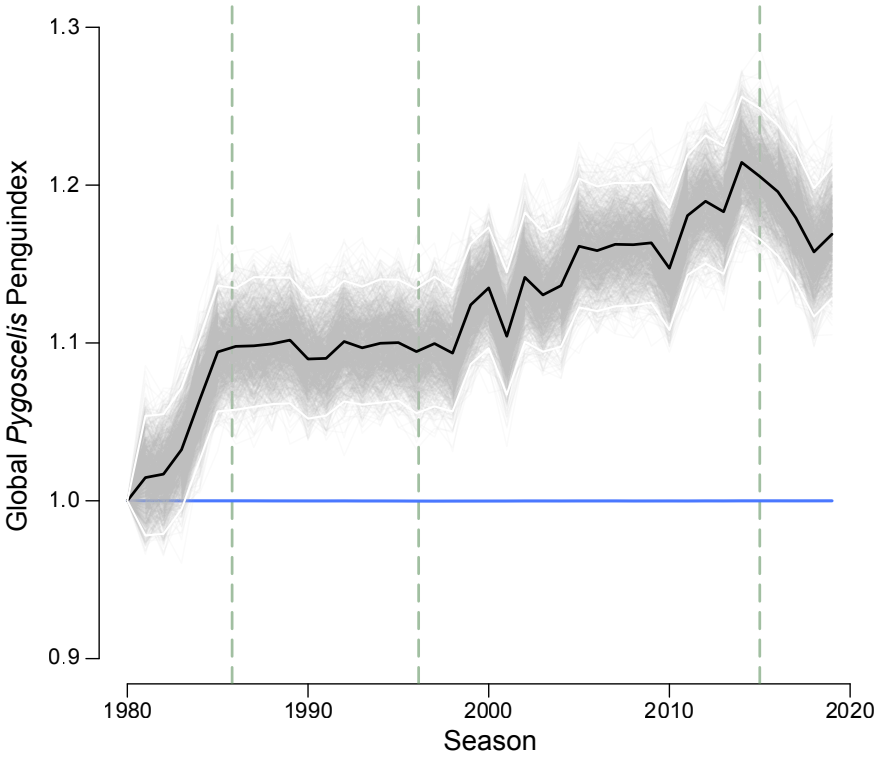

FIGURE 6.1: Penguindex for global *Pygoscelis* penguin populations from 1980-2019. The black line denotes the mean, the white lines the 95% credible intervals, and the gray lines each iteration. Identified change points occur in 1986, 1996, and 2015. The blue line denotes the null model index.

6.3.2 Global species-level trends

The global *Pygoscelis* Penguindex can be disaggregated by species to identify species-specific trends. On average, Adélie penguin populations were mostly stable globally between 1980 and 2019 (Figure 6.2a). Following the global trend for all pygoscelid species, Adélie penguin population abundance increased between 1980-1986, with the index maximum around this time representing a 5.8% increase in the average population (95% CI = 4.0-7.6%). Between 1986 and 1990, the global Adélie index declined quickly (1990 index 1.031, 95% CI = 1.014-1.049), followed by a period of relative stability until 2005 (2005 index 1.035, 95% CI = 1.019-1.050). This was followed by another period of rapid decline until 2010 during which the average Adélie penguin population returned to the 1980 baseline (2010 index 0.997, 95% CI = 0.981-1.013). Between

2010 and 2019, the global Adélie Penguindex remained approximately stable (2019 index 0.994, 95% CI = 0.975-1.014).

Global Chinstrap penguin populations decreased on average by 21.4% (95% CI = 17.1-25.8%) between 1980 and 2019 (Figure 6.2b). While Chinstrap penguin populations declined the most on average globally, declines were not constant, with trends in the global Chinstrap Penguindex displaying the largest number of distinct eras of change compared with the other two *Pygoscelis* species. An initial period of slow growth between 1980-1983 (1983 index 1.015, 95% CI = 0.971-1.060) was followed by one of sharper growth until 1985 (1985 index 1.100, 95% CI = 1.051-1.152). Between 1985 and 1991, the global Chinstrap index declined back to baseline and then fell below baseline (1991 index 0.988, 95% CI = 0.943-1.038). Two eras of slower decline were identified between 1991 and 2013 (2013 index 0.843, 95% CI = 0.800-0.889). Between the 2013 and 2014 seasons, global Chinstrap penguin populations increased by 3% of baseline (2014 index 0.867, 95% CI = 0.823-0.910), before declining quickly to their lowest point in 2019 (2019 index 0.785, 95% CI = 0.742-0.829).

Conversely, Gentoo penguin breeding populations increased by 105.0% (95% CI 87.2-125.5%) globally between 1980-2019 (Figure 6.2c). Prior to 2001, the average Gentoo penguin population increased by 46.6% (95% CI = 43.5-60.7%). This period of growth was followed by a shorter period of even more rapid growth between 2001-2015 (2015 index 2.067, 95% CI = 1.894-2.250). Notably, however, the average Gentoo penguin population was almost completely stagnant between 2015-2019 (2019 index 2.050, 95% CI = 1.872-2.255). Null models for all species-level indices remained stable at 1.0 through 2019.

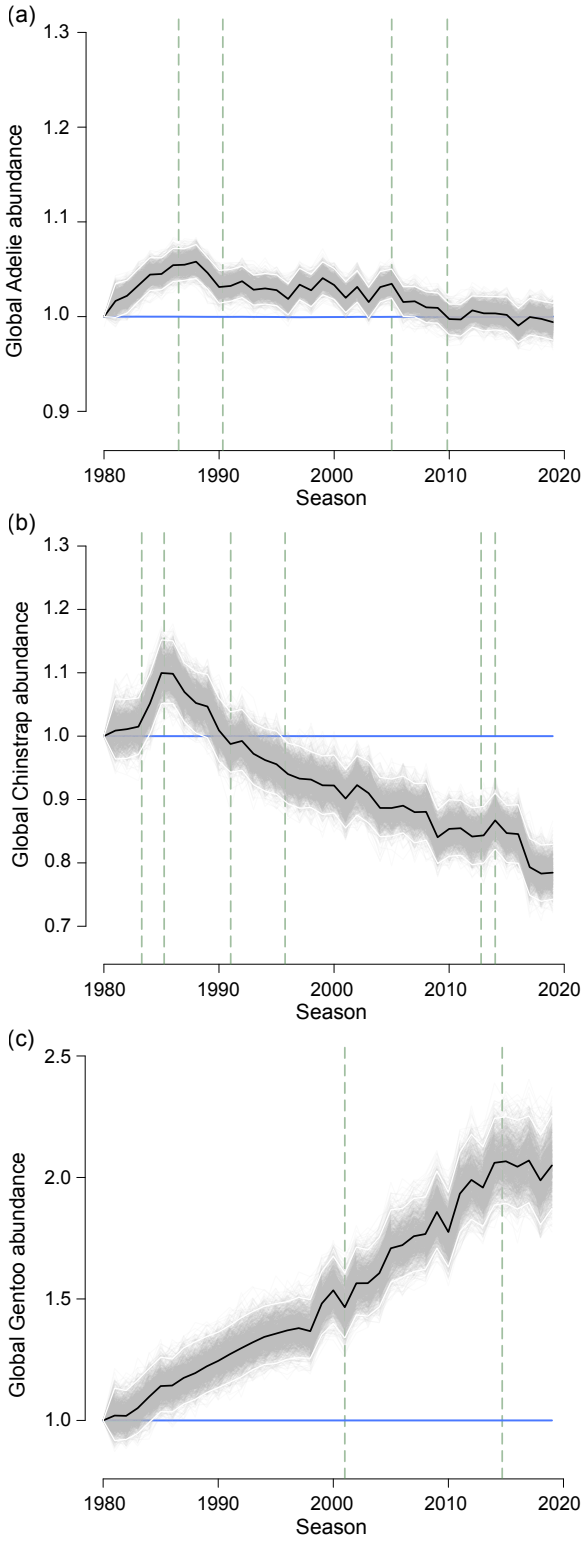

FIGURE 6.2: Species-level Penguindex for global (a) Adélie, (b) Chinstrap, and (c) Gentoo penguin populations from 1980-2019. Each black line denotes the mean, the white lines the 95% credible intervals, and the gray lines each iteration. Each blue line denotes the null model index. Identified change points are reported in Appendix S2: Table S1.

6.3.3 Species-specific regional trends

Regional Adélie trends

Species trends disaggregated by region show differing patterns across the Antarctic (Figures 6.3 and 6.4). Note in Figures 6.3 and 6.4 we choose a layout to emphasize differences between species; a geographic comparison for each species is presented in Appendix S2: Figures S1-S3. The Ross Sea (Region 5) contains the largest number of Adélie penguin breeding colonies (37 populations; Figure 6.3e). Change point analysis identified three major eras of change for these populations, with an initial surge of growth between 1980-1987 leading to an average increase in abundance of 80.8% (95% CI = 57.7-107.0%) of the 1980 baseline. This was followed by a shorter period of rapid decline until 1990 (1990 index 1.098, 95% CI = 0.951-1.259). The period from 1990-2019 displayed large fluctuations, with the overall trend being positive (2019 index 1.510, 95% CI = 1.252-1.789; null model index 1.019). Adélie populations in Eastern Antarctica also increased on average between 1980 and 2019. In Northeastern Antarctica (12 populations; Region 7, Figure 6.4b), Adélie populations increased by 291.9% (95% CI = 106.3-576.4%; the null model increased by 4.9%) on average between 1980-2019, though a period of rapid decline was observed between 2004-2010 (2004 index 4.276, 95% CI 2.532-6.464; 2010 index 2.676, 95% CI 1.624-4.151). Adélie populations in Southeastern Antarctica (19 populations; Region 8, Figure 6.4c) increased on average by 164.8% (95% CI = 70.7-296.9%; the null model increased by 2.6%) over the 40-year time series, though the average population peaked at 472.1% (95% CI 308.1-678.2%) of the 1980 baseline in 2004 before declining rapidly between 2004-2009 (2009 index 2.67, 95% CI 1.876-3.670). Between 2009-2019, Adélie populations in Southeastern Antarctica remained approximately stable on average.

Adélie breeding populations on Elephant Island, South Orkney Islands, and South Shetland Islands (19 populations; Region 3, Figure 6.3c) declined on average 76.2% between 1980-2019 (95% CI = 63.7-85.5%; the null model increased by 7.1%) after an