FIGURE C.4: Prior posterior overlap for standard deviation σ_{R_i} of the Student's t distribution for abundance for ACBR R_i.
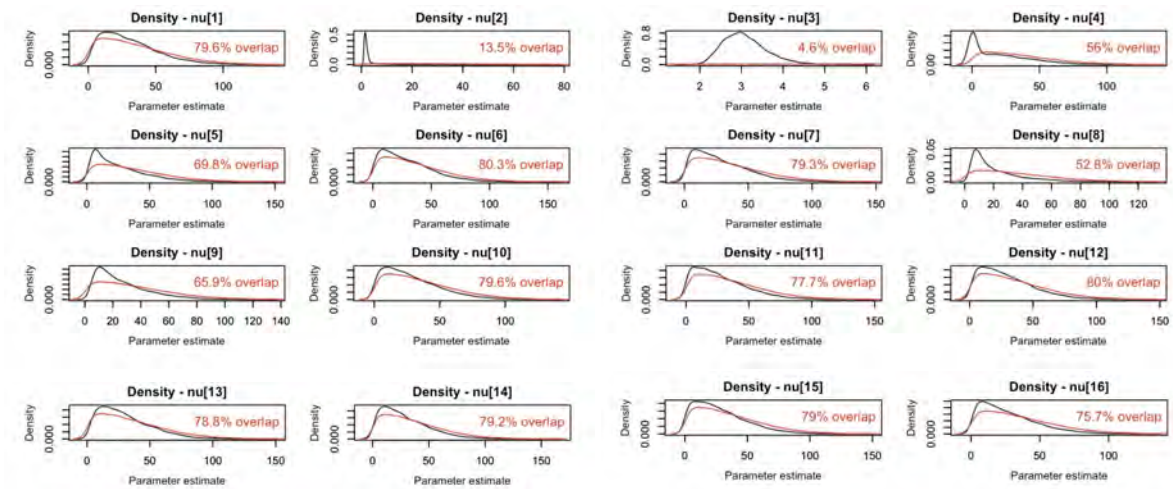
FIGURE C.5: Prior posterior overlap for degrees of freedom ν_{R_i} of the Student's t distribution for abundance for ACBR R_i.

C.3 Phytoplankton growth model

This document describes a version of the Community Earth System Model version 2 (CESM2) ocean ecosystem model, the Marine Biogeochemical Library (MARBL), with four phytoplankton functional types and four zooplankton functional types (hereinafter referred to as "4P4Z"). A thorough documentation for MARBL can be found in [137]; the standard version of MARBL has three phytoplankton types and one zooplankton. Here, trophic relationships and the parameters and equations governing phytoplankton and zooplankton growth are described for the 4P4Z version of the MARBL ecosystem.

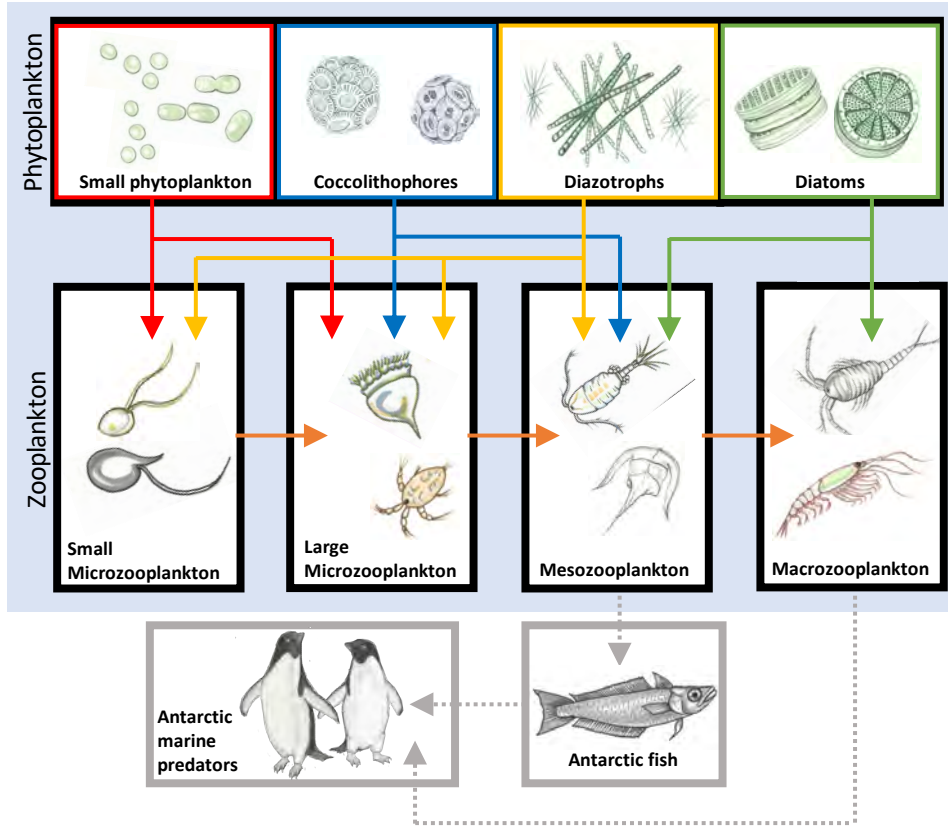## C.3.1 Phytoplankton growth equations and parameterizations

FIGURE C.6: The new 4P4Z MARBL ecosystem configuration in CESM with illustrations of examples of organisms falling into each functional group. CESM components are shown in the light blue background; grazing relationships implemented in the model are indicated by colored arrows. Components outlined in gray show potential connections to higher trophic levels, but are not explicitly simulated in CESM.

The four phytoplankton functional groups simulated by the 4P4Z version of MARBL are small phytoplankton, diatoms, coccolithophores and diazotrophs (Figure C.6, top row). The bulk of marine production comes from small phytoplankton and diatoms, with smaller fractions from the coccolithophores, i.e., calcifying phytoplankton, and nitrogen-fixing phytoplankton, diazotrophs. Phytoplankton growth rates (μ) are parameterized as the product of the resource-unlimited growth rate (μ_{ref}) at a reference temperature (30°C). Diatoms have the fastest μ_{ref}, while diazotrophs have the slowest μ_{ref} (Table C.2). μ_{ref} is scaled by temperature (T_{lim}), nutrient (N_{lim}), and irradiance (I_{lim}) limitation terms:

$$\mu = \mu_{ref} T_{lim} N_{lim} I_{lim}. \tag{C.11}$$

All phytoplankton have the same temperature limitation function, parameterized as a Q10 (temperature coefficient) curve:

$$T_{lim} = Q10^{\left(\frac{T-30°\text{C}}{10°\text{C}}\right)}, \tag{C.12}$$

where $Q10 = 1.7$ in CESM2.

Light limitation (I_{lim}) of phytoplankton is described by the following equation originally from [77]; parameters for each phytoplankton type in the following equation can be found in Table C.2 :

$$I_{lim} = 1 - \exp\left(\frac{-\alpha^{Chl}\theta^C I}{\mu_T N_{lim}}\right), \tag{C.13}$$

where α^{Chl} is the initial slope of the chlorophyll-*a* specific photosynthesis-irradiance (PI) curve and θ^C is the chlorophyll to carbon ratio, *I* is irradiance as photosynthetically active radiation (PAR; W m^{-2}), and N_{lim} is the fractional nutrient limitation term for the most limiting nutrient (see below). Diatoms have the highest θ^C, which provides a competitive advantage under low light conditions.

Nutrient limitation in CESM is represented via Michaelis-Menten uptake kinetics:

$$N_{lim} = \frac{\text{N}}{\text{N} + K_m^{\text{N}}} \tag{C.14}$$

where N represents the ambient concentration of a nutrient and K_m^{N} is the half saturation constant for that nutrient (or aqueous CO_2 to represent carbon limitation in coccolithophores; [120]). Half saturation constants for nutrient uptake for each phytoplankton type are listed in Table C.2. The nutrient which has the lowest N_{lim} is considered the most limiting and is used as the nutrient limitation term in Equation C.11. Small phytoplankton have the lowest K_m^{N} values, conferring a competitive advantage under low nutrient conditions.

Phytoplankton losses include a linear mortality rate, a non-linear aggregation rate, and grazing. Mortality and aggregation rates and formulas are shown in [137] (coccolithophores have the same parameters for these losses as the small phytoplanton group), while zooplankton grazing is described in the following sections.

Parameter	Unit	Definition	cocco	diat	sp	diaz
μ_{max}	d^{-1}	Max. C-specific growth rate	4.7	5.3	4.8	1.9
α	$\dfrac{mmolCm^2}{mgChlWs}$	Initial slope of photosynthesis-irradiance curve	0.30	0.39	0.35	0.31
Θ_{max}	$\dfrac{mgChl}{mmolN}$	Maximum Chl:N ratio	3.5	4.0	2.5	2.5
K_{Fe}	μmol m^{-3}	Fe half saturation constant	0.05	0.05	0.03	0.045
K_{PO4}	mmol m^{-3}	PO_4 half saturation constant	0.006	0.05	0.01	0.015
K_{DOP}	mmol m^{-3}	DOP half saturation constant	0.25	0.5	0.3	0.1
K_{NO3}	mmol m^{-3}	NO_3 half saturation constant	0.41	0.4	0.2	2
K_{NH4}	mmol m^{-3}	NH_4 half saturation constant	0.01	0.05	0.01	0.2
K_{SiO3}	mmol m^{-3}	SiO_3 half saturation constant	n/a	1.6	n/a	n/a
K_{CO2}	mmol m^{-3}	CO_2 half saturation constant	1	n/a	n/a	n/a

TABLE C.2: List of relevant phytoplankton parameterizations used in 4P4Z CESM2-MARBL ecosystem. Abbreviations: Small phytoplankton (sp), diatoms (diat), diazotrophs (diaz), coccolithophores (cocco), dissolved organic phosphorus (DOP).

C.3.2 Zooplankton characteristics and parameterizations

The zooplankton groups are nominally divided up by size, mortality rate (i.e., lifespan or turnover time), and feeding preferences. While we provide some examples of what species these groups could represent (e.g., see Figure C.6, middle row), these are not meant to be specifically describing any particular zooplankton species in great detail. Since this is a global model, what species each zooplankton group represents would depend on geographic location; for instance, the macrozooplankton and mesozooplankton groups could represent Antarctic krill and copepods in the Southern Ocean, respectively. We do not, however, represent multiple life cycle stages within a group or vertical migration behavior. Rather, this model configuration provides an estimate of trophic energy transfer within the lower ranks of the marine food web in an Earth System Model.

General zooplankton characteristics are shown in Table C.3. Gross growth efficiency (GGE), the ratio of growth over ingestion [235], is highest at 0.35 for small microzooplankton in the 4P4Z-MARBL; all other zooplankton groups have a GGE of 0.25, which is within the range of observations [208]. Maximum grazing rates increase as zooplankton size decreases, in line with observations [89]. Linear mortality

rates ($mort$) reflect zooplankton longevity, with faster mortality rates for shorter–lived (smaller) zooplankton groups. A non-linear, density dependent mortality term ($mort2$) represents predation by organisms not explicitly simulated in the model; therefore, the macrozooplankton group has the highest $mort2$ since this group is not preyed upon by the other (smaller) zooplankton in CESM (Table C.3). Zooplankton grazing and mortality rates are scaled by the same Q10 function (Equation C.12) as for phytoplankton above, with a reference temperature (T_{ref}) of 30° and a Q10 of 1.7 (see [137] for more details).

Parameter	Unit	Definition	small microzoo	microzoo	mesozoo	macrozoo
Maximum grazing rate	d^{-1}	sum of g_{max} of all prey at T_{ref}	4.3	2.75	0.25	0.25
GGE	fraction	Gross growth efficiency	0.35	0.25	0.25	0.25
K^P	$mmol/m^3$	Range of half saturation constants for prey	1.3–1.35	1.3	0.75–1.6	0.6
$mort$	d^{-1}	Linear mortality rate at T_{ref}	0.07	0.06	0.0037	0.00125
$mort2$	d^{-1} / $(mmol/m^3)$	Non-linear, density dependent mortality rate at T_{ref}	0.0028	0.0038	0.0048	0.03

TABLE C.3: List of relevant zooplankton parameterizations used in the 4P4Z CESM2-MARBL ecosystem. Abbreviations: small microzooplankton (small microzoo), microzooplankton (microzoo), mesozooplankton (mesozoo), macrozooplankton (macrozoo), reference temperature (T_{ref}, 30°), and maximum grazing rate (g_{max}).

C.3.3   Grazing equations and trophic relationships in 4P4Z MARBL

Grazers ↓	Prey items						
	sp	diaz	cocco	diat	small microzoo	microzoo	mesozoo
small microzoo	4.0	0.3	–	–	–	–	–
microzoo	0.3	0.25	1.3	–	0.9		
mesozoo	–	0.9	0.2	0.8		0.04	
macrozoo				0.2			0.04

TABLE C.4: Maximum grazing rates (g_{max}; d^{-1}) for each grazer-prey relationship in the 4P4Z MARBL. Abbreviations: Small phytoplankton (sp), diatoms (diat), diazotrophs (diaz), coccolithophores (cocco), small microzooplankton (small microzoo), microzooplankton (microzoo), mesozooplankton (mesozoo), macrozooplankton (macrozoo)

Grazing relationships are indicated in Figure C.6, as well as in Tables C.4 and C.5. The small microzooplankton group grazes on small phytoplankton (representing microzooplankton grazing on *Prochlorococcus* and *Synechococcus*; [20]) and diazotrophs (representing microzooplankton grazing on small, unicellular diazotrophs; [232]). Grazing of small phytoplankton by small microzooplankton follows a Holling Type III relationship following findings of [20]:

$$G = g_{max} \cdot T_{lim} \cdot Z \cdot \left(\frac{P^2}{P^2 + K^2} \right) \tag{C.15}$$

where g_{max} is the maximum grazing rate of small microzooplankton (Table C.4), T_{lim} is the same Q10 function as described above for phytoplankton (Equation 1), Z is the zooplankton concentration, P is the small phytoplankton concentration, and K is the half-saturation constant for grazing for this relationship (see Table C.5). All the other grazing relationships in the 4P4Z MARBL configuration are a Holling type II, or Michaelis-Menten type (same variable definitions as for Equation C.15):

$$G = g_{max} \cdot T_{lim} \cdot Z \cdot \left(\frac{P}{P + K} \right) \tag{C.16}$$

The microzooplankton group in the 4P4Z MARBL has a diverse diet consisting of small phytoplankton, diazotrophs, coccolithophores, as well as small microzooplankton (Figure C.6). The mesozooplankton diet in the model consists primarily of diatoms, reflecting observations that diatoms serve as the dominant food source for copepods [105]. The mesozooplankton group also grazes on coccolithophores, diazotrophs (e.g., Trichodesmium), and microzooplankton. The macrozooplankton group also grazes primarily on diatoms (e.g., krill grazing on diatoms; [86]), with lower grazing rates on the mesozooplankton group. Diatoms have lower overall grazing rates than other phytoplankton in the model due to protection from grazing, e.g., see [201]. Coccolithophores are grazed by both microzooplankton and mesozooplankton, reflecting the wide size range of this group [15, 170].

	Prey items						
Grazers ↓	sp	diaz	cocco	diat	small microzoo	microzoo	mesozoo
small microzoo	1.35	1.3	–	–	–	–	–
microzoo	1.3	1.3	1.3	–	1.3		
mesozoo	–	0.8	1.1	0.75		1.6	
macrozoo				0.6			0.6

TABLE C.5: Half saturation constants for grazing (K^P; mmol m^{-3}) for each grazer-prey relationship in the 4P4Z MARBL. Abbreviations: Small phytoplankton (sp), diatoms (diat), diazotrophs (diaz), coccolithophores (cocco), small microzooplankton (small microzoo), microzooplankton (microzoo), mesozooplankton (mesozoo), macrozooplankton (macrozoo)

C.3.4 Model validation and further details

Model evaluation using oceanic observational databases and a complete list of parameters is archived on Zenodo: `https://doi.org/10.5281/zenodo.7191370` [216].

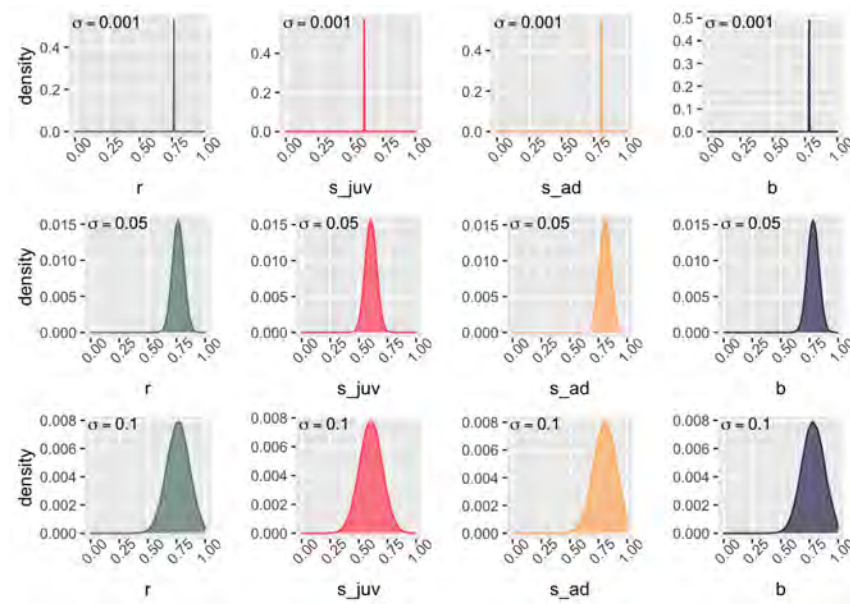# C.4 Age-structured simulations: Gaussian inputs

FIGURE C.7: Density plots for all four demographic parameters (by column: r, reproductive success; s_{juv}, juvenile survival; s_{ad}, adult survival; b, breeding propensity). Truncated normal distributions are shown centered about the mean values for each parameter (equal to the posterior medians from the ABC parameter estimation: $\mu_r = 0.765$, $\mu_{s_{juv}} = 0.605$, $\mu_{s_{ad}} = 0.808$, $\mu_b = 0.782$) with standard deviations (by row) $\sigma = 0.001, 0.05, 0.1$ In the simulation study, standard deviations for each parameter vary from $0.001 - 0.1$ (with $n = 15$ steps); thus shown here are the minimum variation (top row), mid-range variation (middle row), and maximum variation (bottom row) for each parameter.
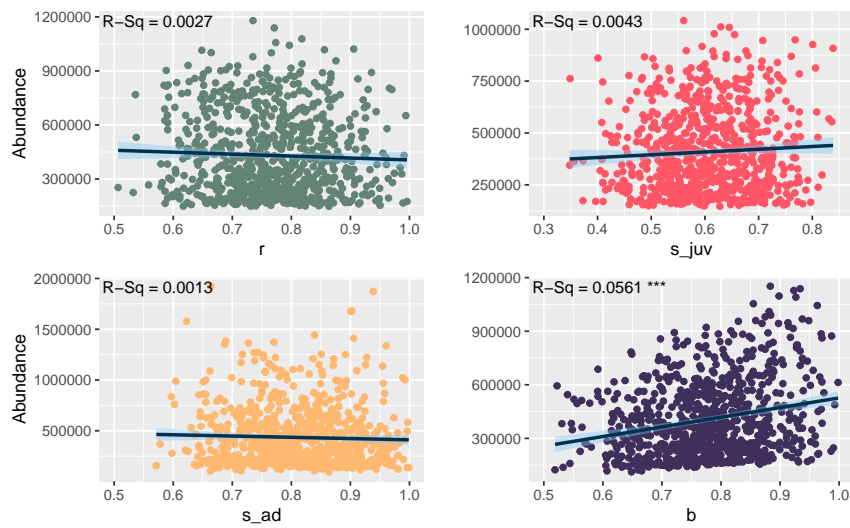
FIGURE C.8: Scatter plots of breeding abundance in year t vs. each of the four demographic parameters in year t. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
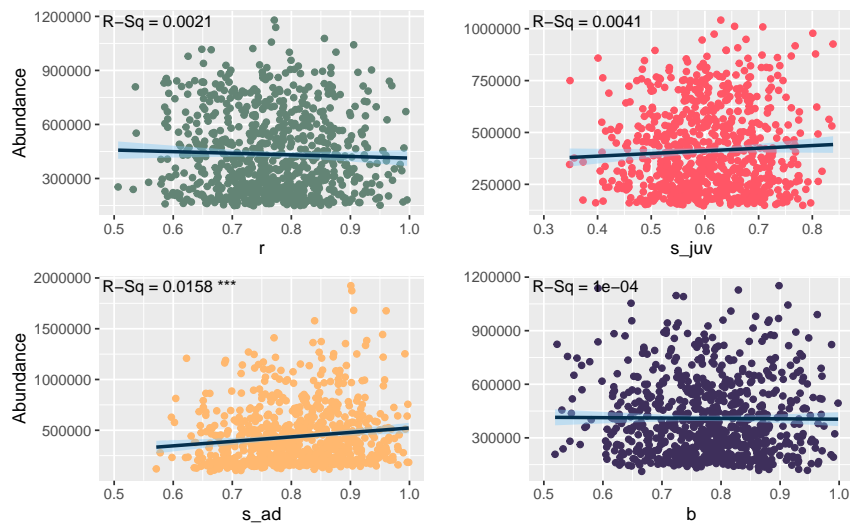

FIGURE C.9: Scatter plots of breeding abundance in year t vs. each of the four demographic parameters in year $t - 1$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
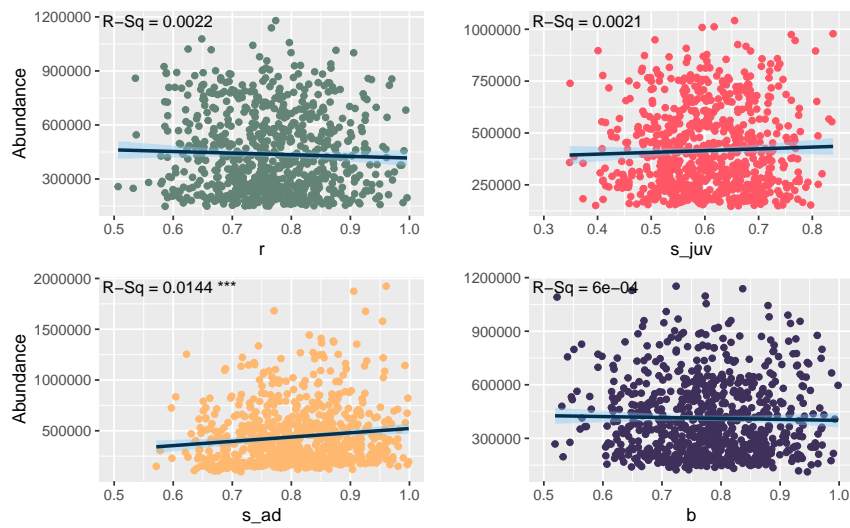
FIGURE C.10: Scatter plots of breeding abundance in year t vs. each of the four demographic parameters in year $t-2$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
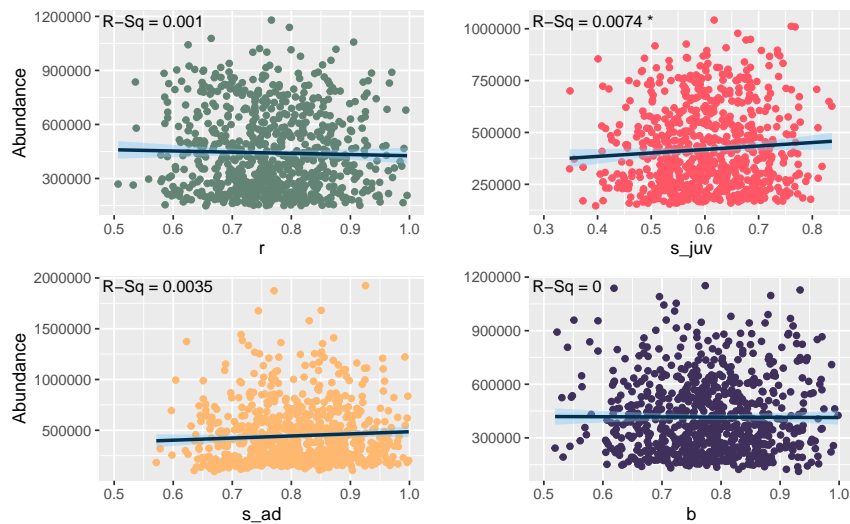

FIGURE C.11: Scatter plots of breeding abundance in year t vs. each of the four demographic parameters in year $t-3$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
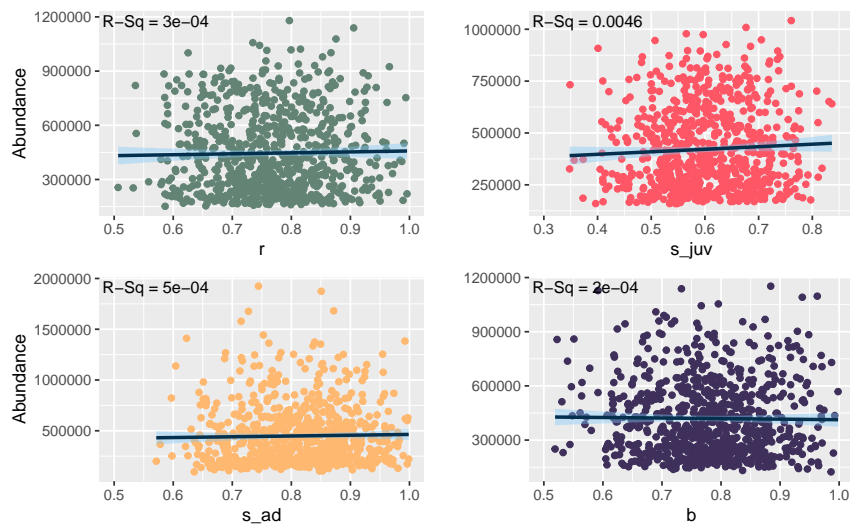
FIGURE C.12: Scatter plots of breeding abundance in year t vs. each of the four demographic parameters in year $t - 4$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.

FIGURE C.13: Scatter plots of breeding abundance in year t vs. each of the four demographic parameters in year $t - 5$. 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 $p < 0.001$.
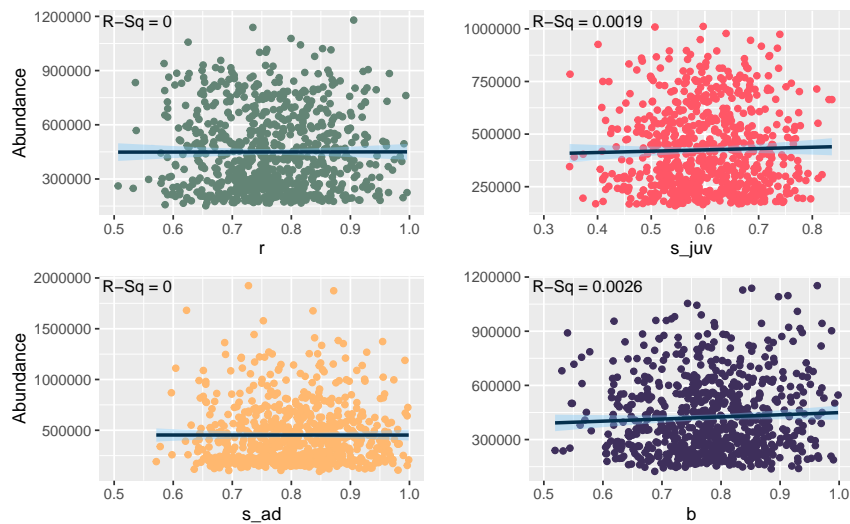
FIGURE C.14: Scatter plots of breeding population growth in year t vs. each of the four demographic parameters in year t. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
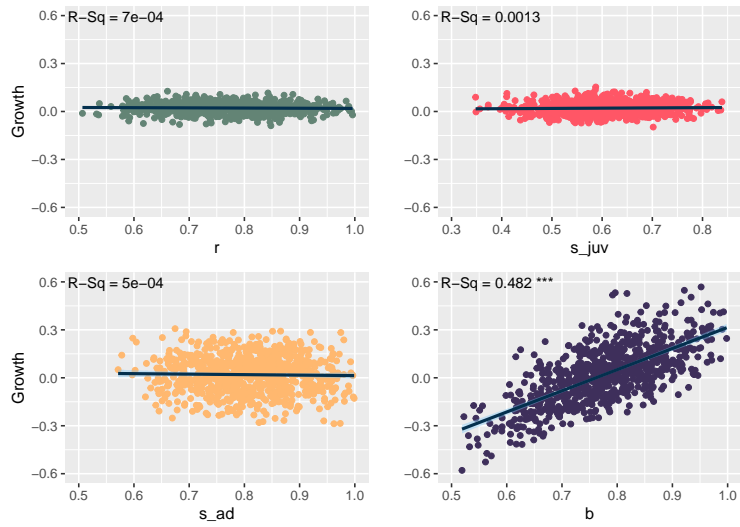

FIGURE C.15: Scatter plots of breeding population growth in year t vs. each of the four demographic parameters in year $t - 1$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
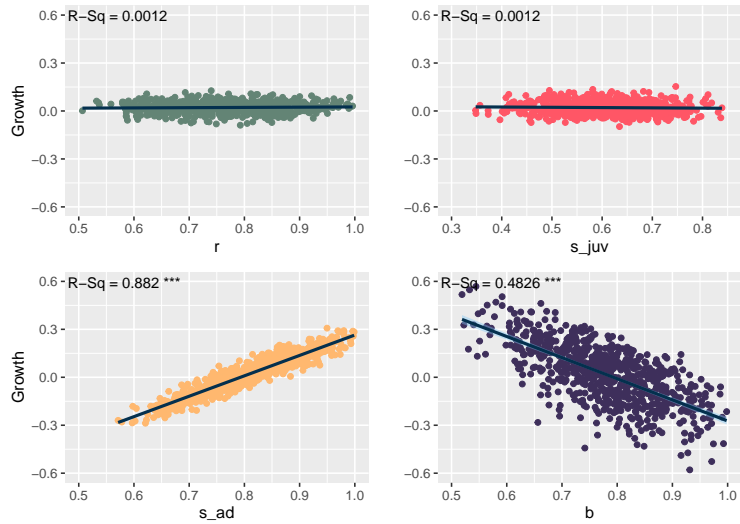
FIGURE C.16: Scatter plots of breeding population growth in year t vs. each of the four demographic parameters in year $t - 2$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
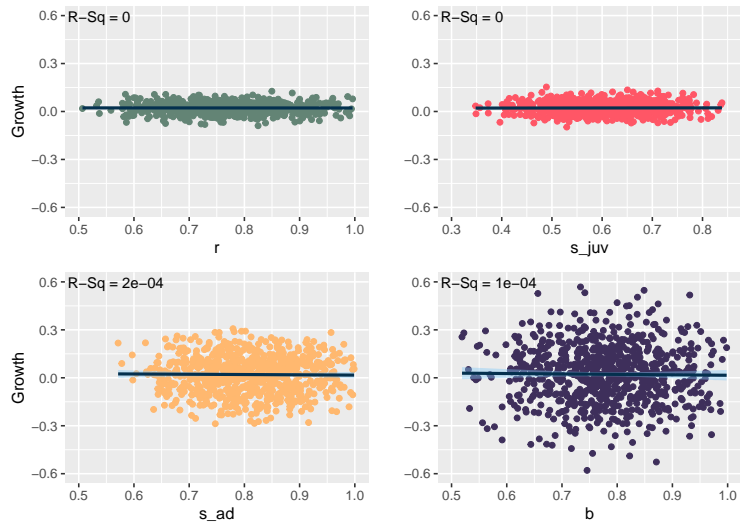

FIGURE C.17: Scatter plots of breeding population growth in year t vs. each of the four demographic parameters in year $t - 3$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.

FIGURE C.18: Scatter plots of breeding population growth in year t vs. each of the four demographic parameters in year $t - 4$. 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 $p < 0.001$.
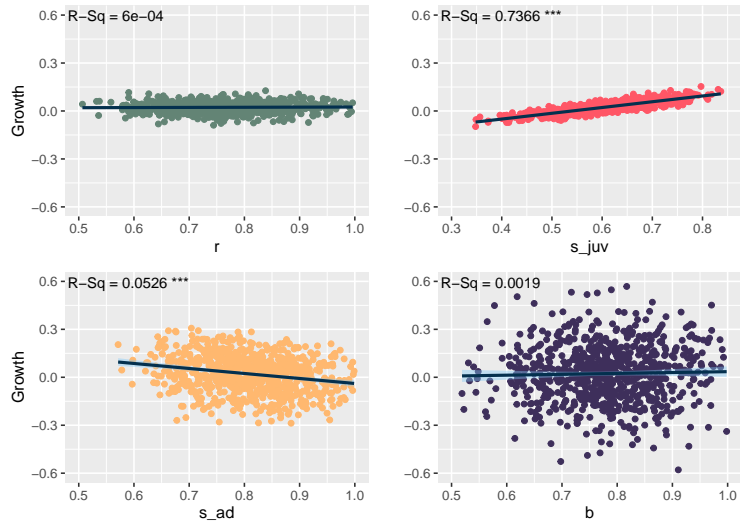

FIGURE C.19: Scatter plots of breeding population growth in year t vs. each of the four demographic parameters in year $t - 5$. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
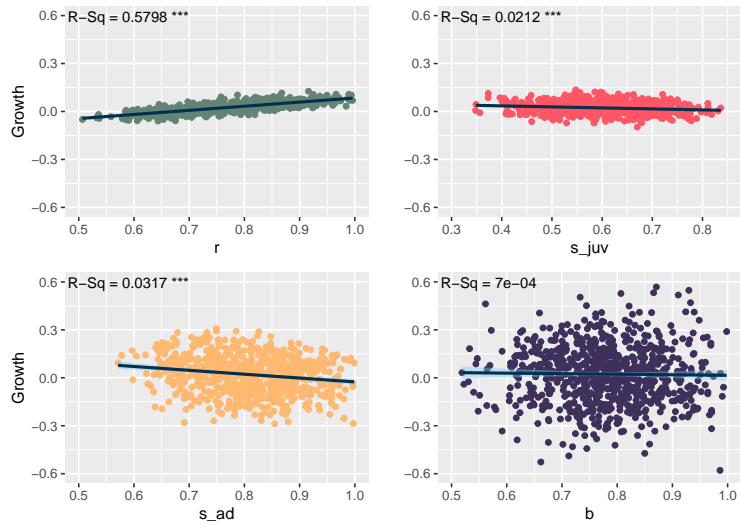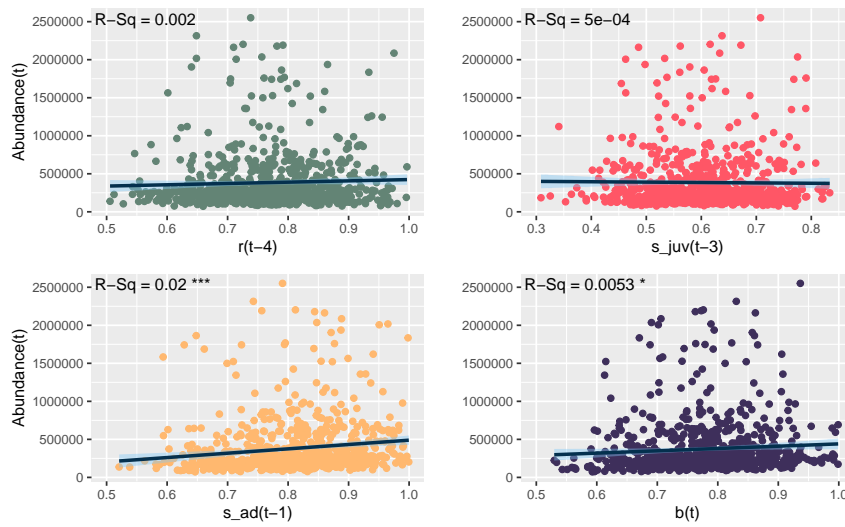
FIGURE C.20: Scatter plots of breeding abundance vs. each of the four demographic parameters with varied lags, with the standard deviations of all parameters being high (0.1). Lags are as follows: 4 years for r, 3 years for s_{juv}, 1 year for s_{ad}, and 0 years for b. Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$ while * indicates $p < 0.05$.
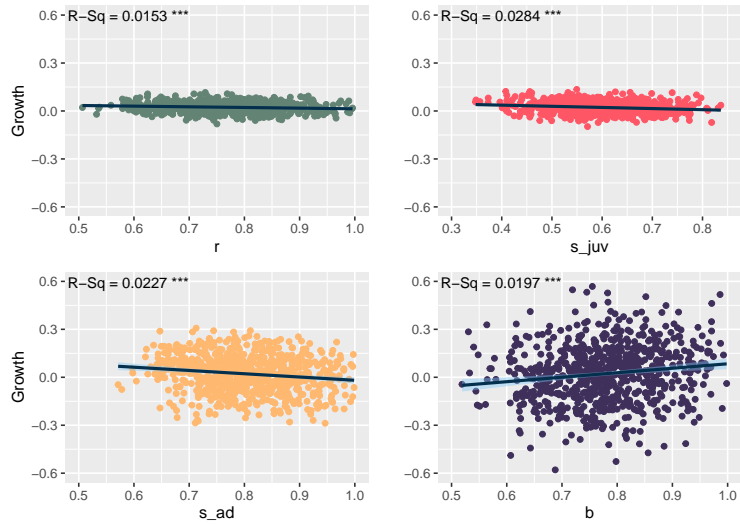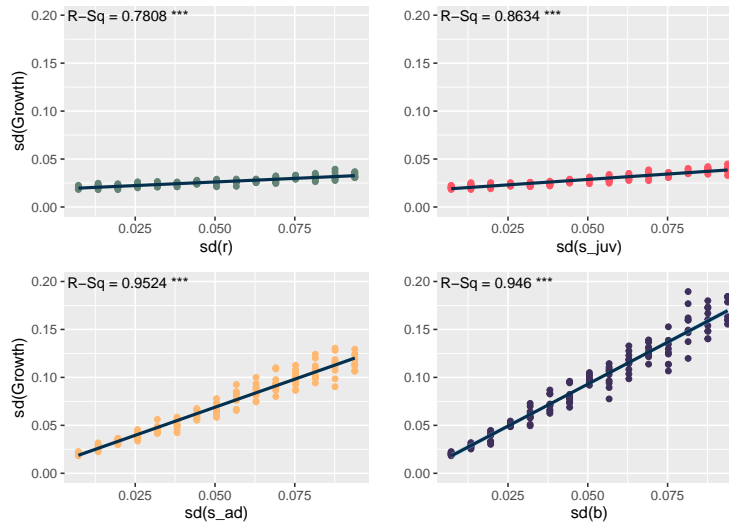

FIGURE C.21: Scatter plots of standard deviation in growth of breeding abundance vs. standard deviation in each of the four demographic parameters. In each plot, the standard deviation of the parameter of interest increases from 0.001 to 0.1 along the x-axis (with ten iterates per value of σ), while the standard deviations of the other three parameters are constant at 0.001.
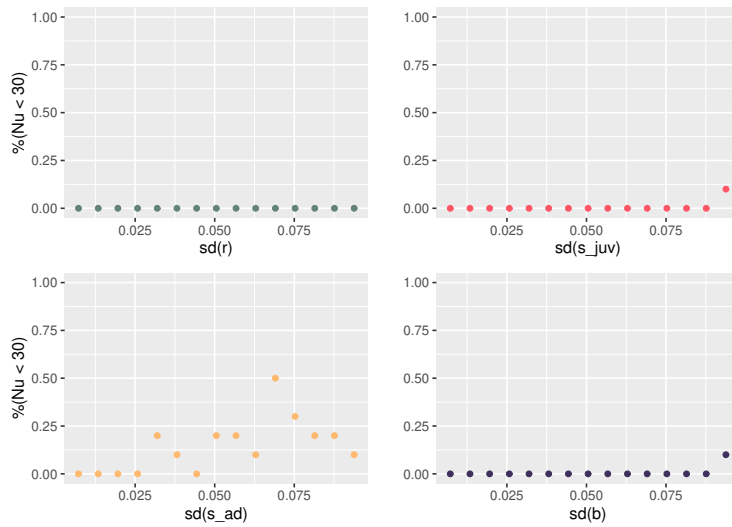
FIGURE C.22: Scatter plots of the percentage of iterates (out of $m = 10$) in which the distribution for abundance was fitted to be heavy-tailed (i.e. $\nu < 30$) vs. the standard deviation of each of the four demographic parameters. In each plot, the standard deviation of the parameter of interest is increasing from $0.001 - 0.1$ along the x-axis, while the standard deviations of the other three parameters are constant at 0.001.
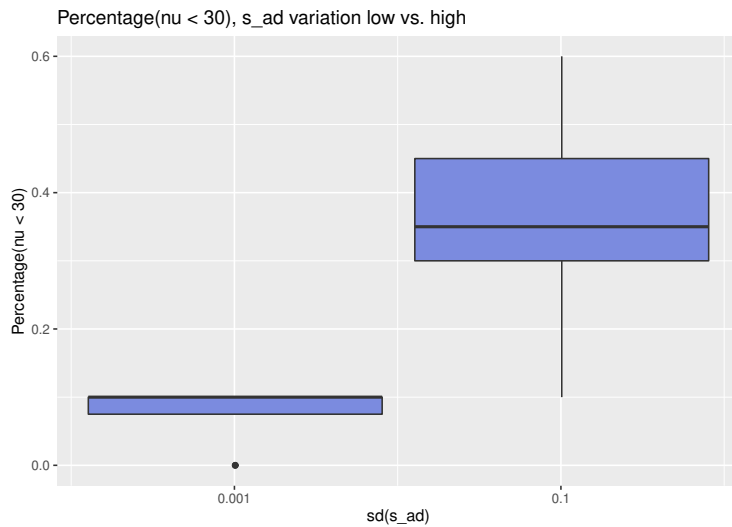
FIGURE C.23: Boxplots for the percentage of time series iterates with heavy-tailed dynamics ($\nu < 30$) for adult survival with low variability ($\sigma = 0.001$, with all other parameters with either $\sigma = 0.001$ or 0.1, sample size of 8 percentages out of 10 iterates) and adult survival with high variability ($\sigma = 0.1$, with all other parameters with either $\sigma = 0.001$ or 0.1, sample size of 8 percentages out of 10 iterates). All parameters are drawn from Gaussian distributions. One-way ANOVA is significant with $p < 0.001$.

184

C.5 Age-structured simulations: Heavy-tailed inputs
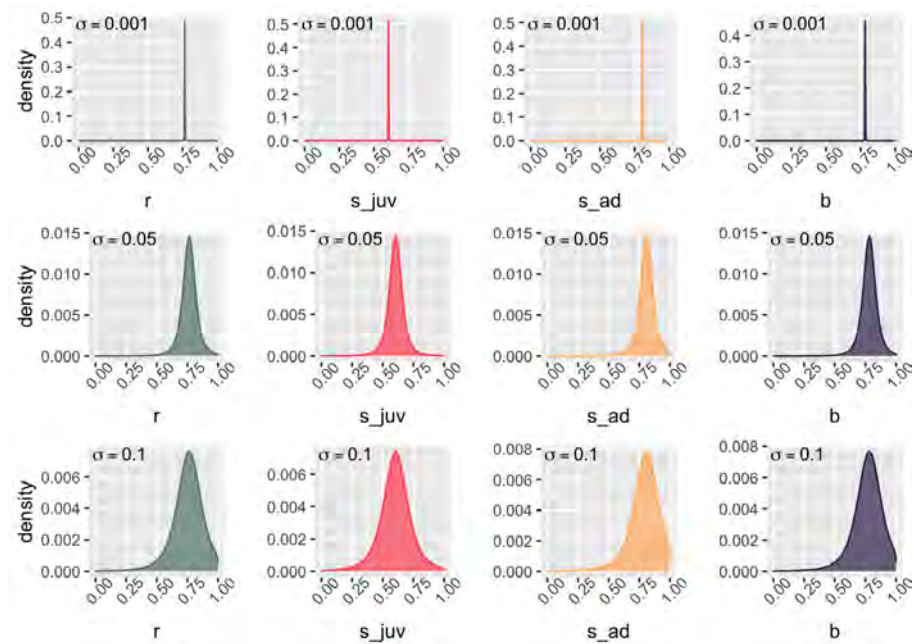
FIGURE C.24: Density plots for all four demographic parameters with heavy-tailed dynamics (by column: r, reproductive success; s_{juv}, juvenile survival; s_{ad}, adult survival; b, breeding propensity). Truncated Student's t distributions are shown centered about the mean values for each parameter (equal to the posterior medians from the ABC parameter estimation: $\mu_r = 0.765$, $\mu_{s_juv} = 0.605$, $\mu_{s_ad} = 0.808$, $\mu_b = 0.782$) with standard deviations (by row) $\sigma = 0.001, 0.05, 0.1$ In the simulation study, standard deviations for each parameter vary from $0.001 - 0.1$ (with $n = 15$ steps); thus shown here are the minimum variation (top row), mid-range variation (middle row), and maximum variation (bottom row) for each parameter.

FIGURE C.25: Scatter plots of breeding abundance vs. each of four demographic parameters with heavy-tailed inputs (i.e. demographic parameters are varying according to a Student's t distribution) 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 $p < 0.001$ while ** indicates $p < 0.01$.
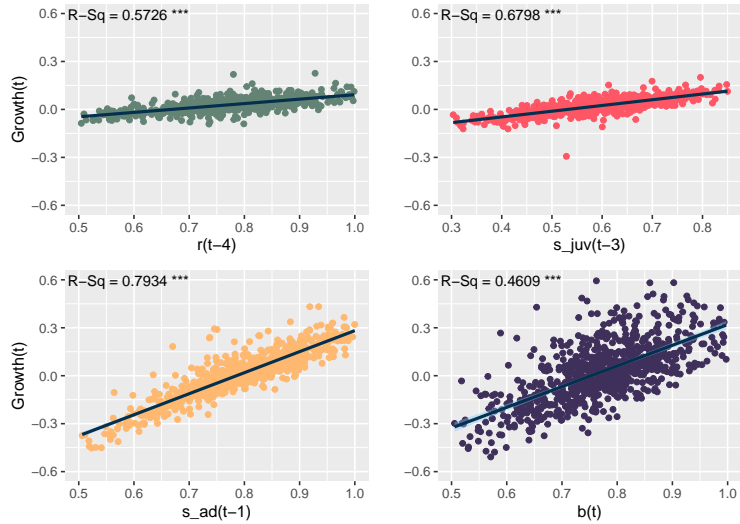
FIGURE C.26: Scatter plots of growth in breeding abundance vs. each of four demographic parameters with heavy-tailed inputs (i.e. demographic parameters are varying according to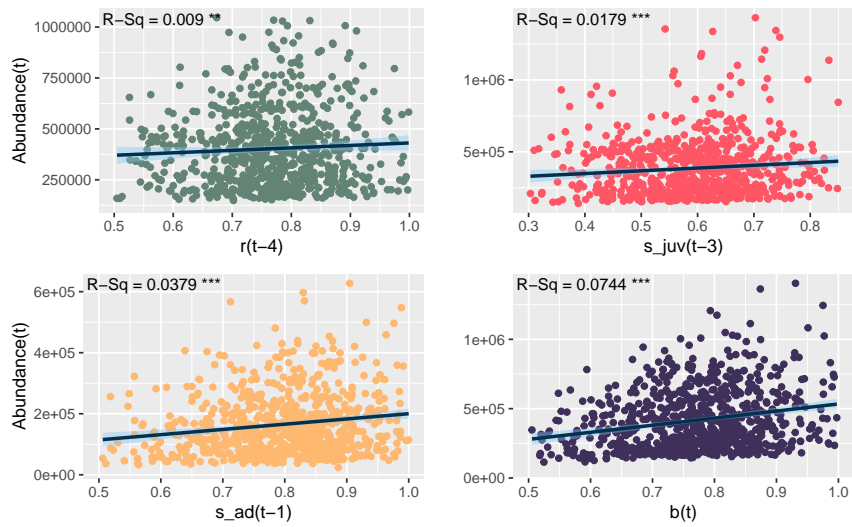 a Student's t distribution) with various lags. Lags are as follows: 4 years for r, 3 years for s_{juv}, 1 year for s_{ad}, and 0 years for b. In each plot, the standard deviation of the parameter of interest is high (0.1), while the standard deviations of the other three parameters are low (0.001). Regression lines (dark blue) and 95% confidence intervals (light blue) are shown, along with R-squared values. *** indicates a significant correlation with $p < 0.001$.
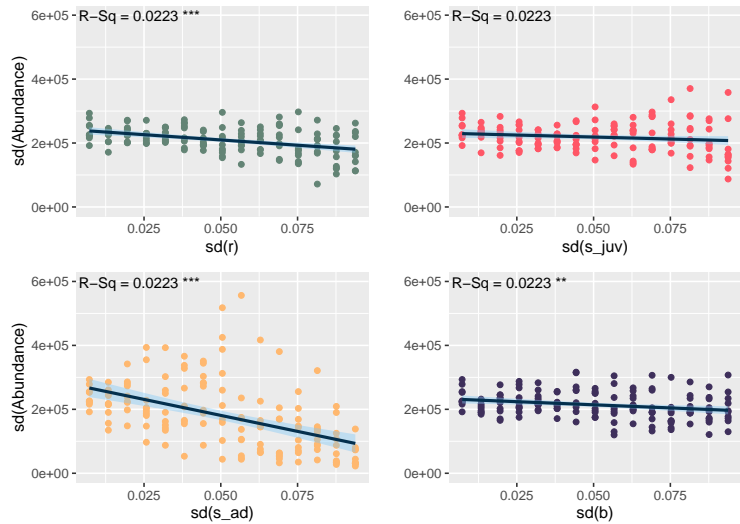
FIGURE C.27: Scatter plots of standard deviation in breeding abundance vs. standard deviation in each of the four demographic parameters, for heavy-tailed inputs. In each plot, the standard deviation of the parameter of interest increases from 0.001 to 0.1 along the x-axis (with ten iterates per value of σ), while the standard deviations of the other three parameters are constant at 0.001. *** indicates a significant correlation with $p < 0.001$ while ** indicates $p < 0.01$.

187
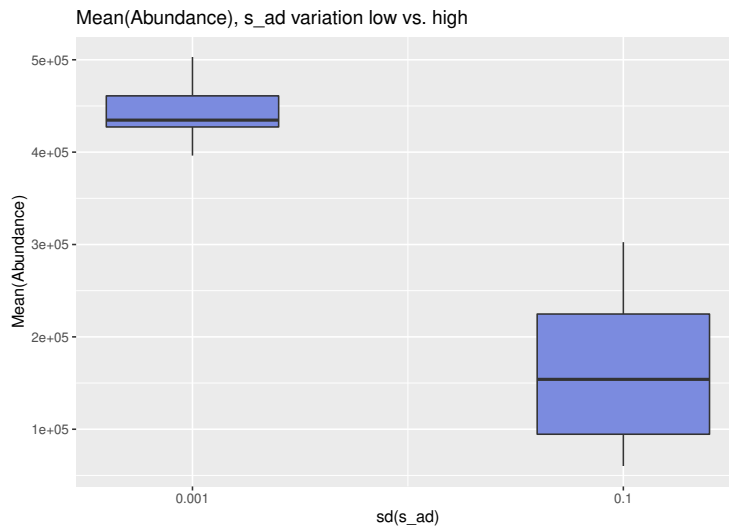
FIGURE C.28: Boxplots for the standard deviation of abundance across time series for heavy-tailed demographic parameters with low variability ($\sigma = 0.001$, sample size of 10 time series iterates) and high variability ($\sigma = 0.1$, sample size of 10 time series iterates). One-way ANOVA is significant with $p < 0.001$.
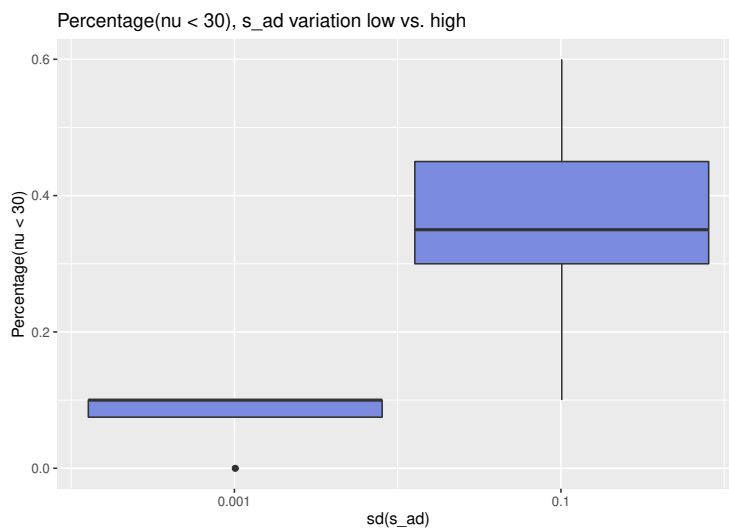
FIGURE C.29: Boxplots for the percentage of time series iterates with heavy-tailed dynamics ($\nu < 30$) for adult survival with low variability ($\sigma = 0.001$, with all other parameters with either $\sigma = 0.001$ or 0.1, sample size of 8 percentages out of 10 iterates) and adult survival with high variability ($\sigma = 0.1$, with all other parameters with either $\sigma = 0.001$ or 0.1, sample size of 8 percentages out of 10 iterates). All parameters are drawn from heavy-tailed ($\nu = 3$) Student's t distributions. One-way ANOVA is significant with $p < 0.01$.

C.6 Age-structured simulations: Comparing Gaussian and heavy-tailed inputs
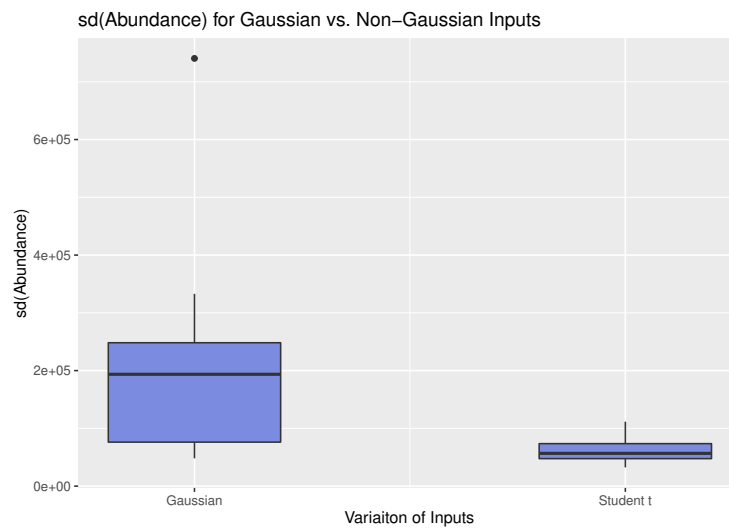
FIGURE C.30: Boxplots for the standard deviation of abundance across time series for demographic parameters with high Gaussian variability ($\sigma = 0.1$) and high heavy-tailed ($\nu = 3$) Student's t variability ($\sigma = 0.1$). One-way ANOVA is significant with $p < 0.01$.

Appendix D

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

D.1 Problems arising from replacing negative simulated abundances with missing data

Consider a simulation framework for generating fluctuating time series similar to that described in Equation (5.5) that replaces negative simulated abundance values (see Figure 5.1b) with missing data (i.e., each negative value is replaced with an 'NA' in R). Therefore the density of the underlying datasets used in the simulations of, for example, 1% population fluctuations and 31% population fluctuations would be vastly different. For populations with high interannual fluctuations, the simulated time series would have far more missing values. This causes several confounding problems in the calculation of the LPI. First, the GAM used to smooth and interpolate population time series may now be fit with fewer data points, likely resulting in a poorer fit. Alternatively, if the removal of negative data points results in a time series with a total of fewer than 6 measured values, that time series will be interpolated using the chain method (log-linear interpolation) instead of the GAM (see Section 5.1.1). Therefore, when population time series have higher levels of interannual fluctuation, the chain method may be used more frequently than for time series with smaller fluctuations and, when the GAM is used to interpolate these time series, it is doing so with less information. In addition, it is important to note that the interpolation of these missing values would only occur when calculating the LPI; thus there is no way to check if the simulated time series are stable on average since interpolation is left to the LPI framework.

Importantly, the decreased data density and the resulting effect on interpolation would only occur for decreasing time series. When population fluctuations cause a population time series to increase on average from the initial population count, the calculation of the LPI would not be affected. Thus the simulated datasets resulting from this simulation framework would be characterized by increasing time series that have data coverage identical to that of the observed population time series and decreasing time series that are more sparse and may be shorter on average. Additionally, since the LPI only considers those population time series that have started but not yet ended in any given year, decreased data density may cause many decreasing population time

series to be removed from the LPI calculations for later years in the range considered. This would result in a higher proportion of population time series included in the calculation of the LPI that are increasing on average, leading to an increasing LPI.

Thus this simulation framework would not only result in similar difficulties generating a suite of time series that are actually stable on average, but it would also result in several additional and confounding issues when calculating the LPI as a result of the inconsistent densities of the underlying datasets. The use of any such simulation framework would be entirely inappropriate to produce a null model for the LPI, and any application of the LPI to datasets with varying data density should be approached with extreme caution.

D.2 Supplementary Results

FIGURE D.1: The LPI for fluctuating, but otherwise stable, populations (mean and interquartile range of 100 iterations shown) drawn from a truncated normal distribution. Population fluctuations are benchmarked to the previous year's population count (see Equation 5.6).
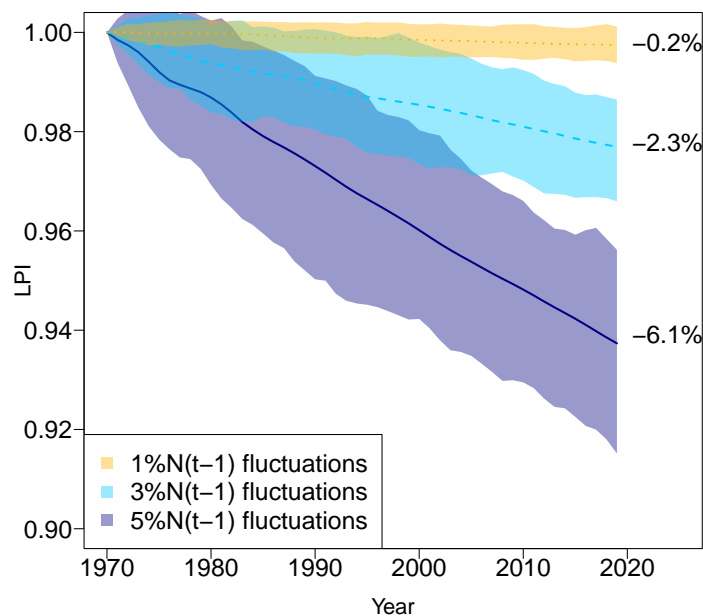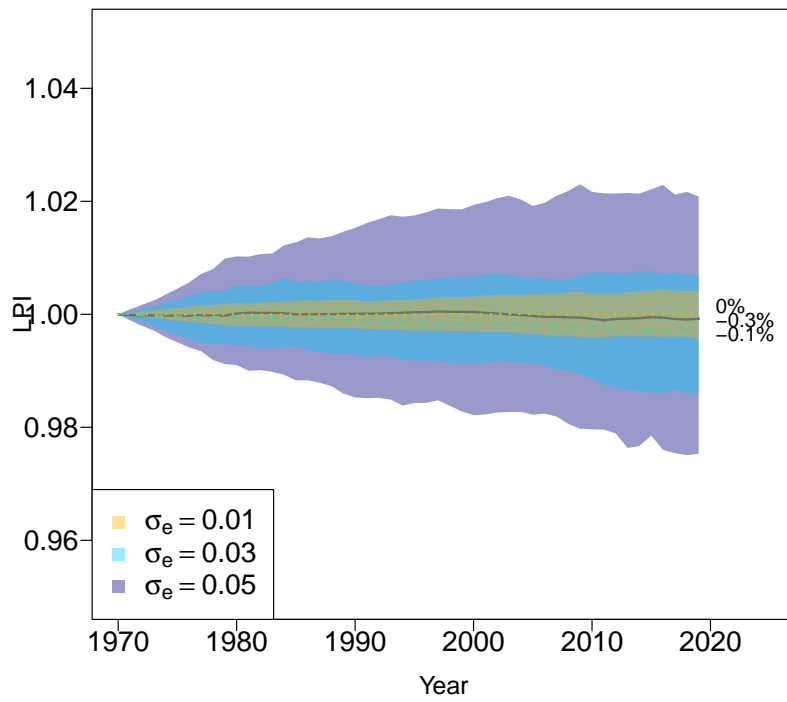
FIGURE D.2: The LPI for fluctuating, but otherwise stable, populations (mean and interquartile range of 100 iterations shown) drawn from a log-normal distribution (see Equation 5.8).

Appendix E

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

E.1 Supplementary Methods

E.1.1 Bayesian state-space models

For each pygoscelid species, we fit a Bayesian state-space model, adapted from [39], to estimate annual penguin nest abundances for each breeding population from 1970 – 2020. 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. As Adélie penguins are spread over a large geographical region spanning the entirety of Antarctica, the Adélie model includes regional effects. We separate three Adélie regions for this model, including breeding sites on the Antarctic Peninsula (CCAMLR subarea 48.1, 48.2), Ross Sea (CCAMLR subarea 88.1, 88.2, 88.3), or Eastern Antarctica (CCAMLR subarea 58.4.1, 58.4.2).

For the purposes of clarity we introduce our nest abundance models in a series of steps, first modeling abundance on the arithmetic scale using the log-normal distribution. For each species, we started by modeling "true" (hereafter latent) nest abundance $z_{i,t}$ at the i_{th} breeding population (or, 'site') in the t_{th} season as:

$$z_{i,t} \sim \text{log-normal}\left(\mu_{i,t} = \log(z_{i,t-1}e^{r_{i,t}}),\ \sigma^2\right) \tag{E.1}$$

where the median of the log-normal distribution, $e^{\mu_{i,t}}$, is a deterministic model for discrete exponential growth, such that nest abundance $z_{i,t}$ is the product of nest abundance in the previous season $z_{i,t-1}$ and the intrinsic rate of growth, $r_{i,t}$. We modeled the intrinsic growth rate for Chinstrap penguins and Gentoo penguins as the sum of a grand mean β, site effects η_i, and seasonal effects ϵ_t as:

$$r_{i,t} = \beta + \eta_i + \epsilon_t \tag{E.2}$$

where each is modeled hierarchically as:

$$\eta_i \sim N\left(0,\ \sigma^2_{site}\right) \tag{E.3}$$

$$\epsilon_t \sim N\left(0,\ \sigma^2_{season}\right). \tag{E.4}$$

193

For Adélie penguins, the intrinsic growth rate for a population in region r is modeled as the sum of a regional mean $\beta_{R[i]}$, site effects η_i, and seasonal effects $\epsilon_{R[i],t}$ (we allow these to differ between the three regions) as:

$$r_{i,t} = \beta_{R[i]} + \eta_i + \epsilon_{R[i],t} \tag{E.5}$$

where each is modeled hierarchically as:

$$\beta_R \sim N\left(0,\, \sigma_{region}^2\right) \tag{E.6}$$

$$\eta_i \sim N\left(0,\, \sigma_{site}^2\right) \tag{E.7}$$

$$\epsilon_{R,t} \sim N\left(0,\, \sigma_{season_R}^2\right). \tag{E.8}$$

Following [39], we then introduced a term to shift these distributions to the left as a corrective for biases that result when summing log-normal distributions [215], leading to a mean-adjusted model for logged latent nest abundance:

$$\log(z_{i,t}) \sim N\left(\mu_{i,t} = \log(z_{i,t-1}) + r_{i,t} - \sigma^2/2,\, \sigma^2\right). \tag{E.9}$$

For breeding populations whose first season of data was later than 1970, we hindcast nest abundances from the first census backwards to 1970. Since the exponential growth function can be inverted, hindcasting nest abundances is functionally no different than forecasting nest abundances into the future or in seasons of missing data within a site's time series. Thus the logged latent nest abundances for the i_{th} site from 1970 to the season of the initial count were modeled as:

$$\log(z_{i,t-1}) \sim N\left(\log(z_{i,t}) - r_{i,t} - \sigma^2/2,\, \sigma^2\right). \tag{E.10}$$

Chick counts were converted to nest counts using an estimate of reproductive success as in [39]. Census data vary with respect to observation error, and counts were provided by the observer(s) along with an associated accuracy score from which observation error can be estimated (see [39]). We modeled the logged observed nest counts y recorded at the i_{th} breeding site in the t_{th} season as:

$$y_{i,t} \sim N\left(\log(z_{i,t}) - \sigma_{i,t}^2/2,\, \sigma_{i,t}^2\right) \tag{E.11}$$

where $\sigma_{i,t}^2$ is the observation error in the recorded count, computed from the accuracy reported by the observer.

For each of the three species modeled, we sampled from six Markov Chain Monte Carlo Markov chains. Each chain was initiated by assuming an initial logged nest abundance drawn from a normal distribution with a mean of 0 and a standard deviation of 1,000,000. In each chain, after an adaptation phase of 3,000 iterations, the first 300,000 iterations were discarded as burn-in. Of the remaining 100,000 iterations, a thinning rate of 100 yielded 1,000 samples from each posterior distribution. We assessed convergence through visual inspection of trace plots and using the Gelman-Rubin diagnostic [79]. We verified that the model adequately captured the dispersion in nest counts (measured as the sum of the squared residuals) using posterior predictive checks and confirmed that the Bayesian p-value (defined as the probability that

the simulated data were more extreme than the observed data) was indicative of a good model fit. These analyses were performed using R v.4.0.2 [179] and JAGS ('Just Another Gibbs Sampler' v.4-10, [177]), interfaced using the `rjags` package (v.4-10, [176]).

Full methodology, code, and results for the each state-space model can be found in the documentation linked below:

Adélie: `https://cchecastaldo.github.io/MAPPPD_adelie_models/index.html`

Chinstrap: `https://cchecastaldo.github.io/MAPPPD_chinstrap_models/index.html`

Gentoo: `https://cchecastaldo.github.io/MAPPPD_gentoo_models/index.html`

E.1.2 Calculating the Penguindex

FIGURE E.1: Map of Antarctica showing Adélie penguin breeding populations that were (a) included and (b) excluded from the Penguindex calculations, each population is colored by region.

We subset the population time series used to calculate the Penguindex as described in the manuscript. Figures E.1-E.3 show the breeding populations that were included and excluded from this calculation for each species. (Note, all breeding population time series were included in the Bayesian state-space model for each species.) The included time series were used to calculate regional, species, and global trends, which were then converted to index values (Figure E.4).

FIGURE E.2: Map of the Antarctic Peninsula showing Chinstrap penguin breeding populations that were (a) included and (b) excluded from the Penguindex calculations, each population is colored by region. Insets show (from top to bottom) the South Sandwich Islands and Balleny Islands.
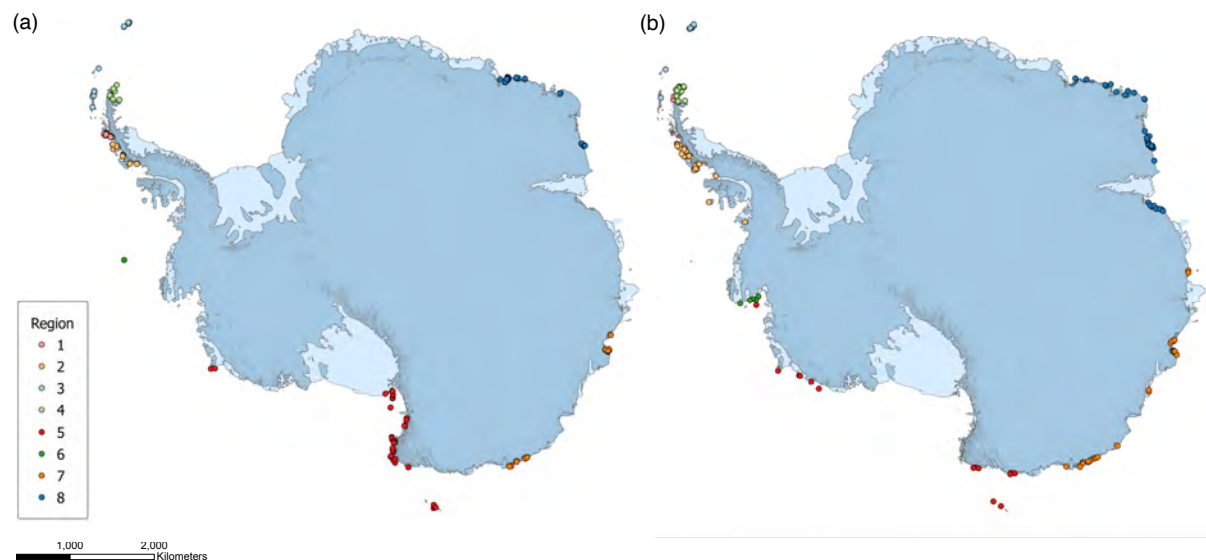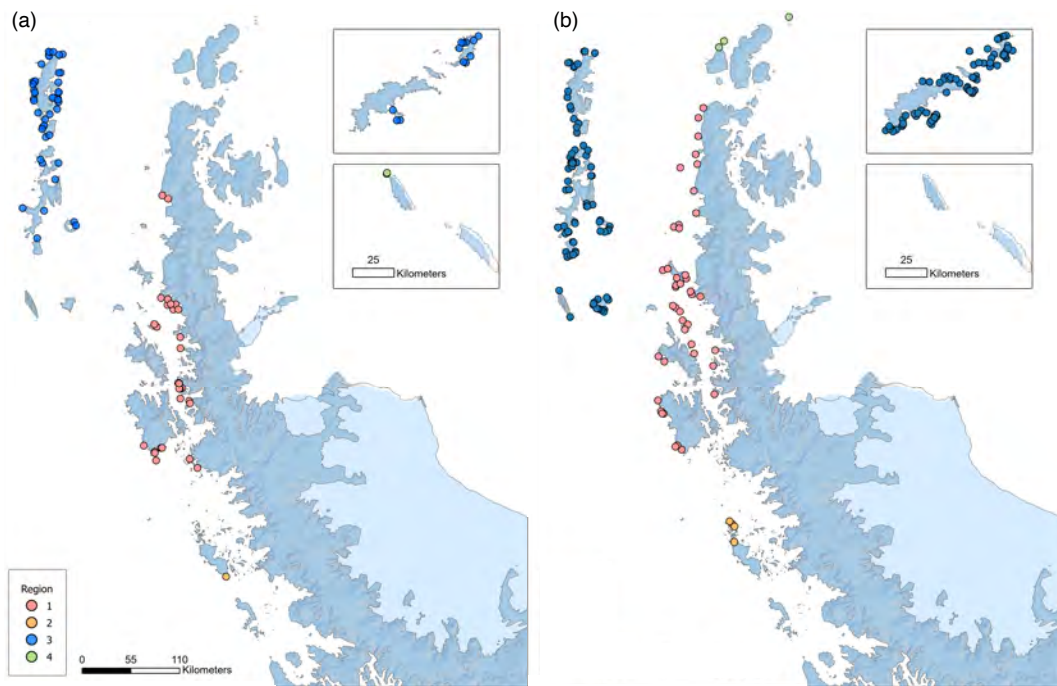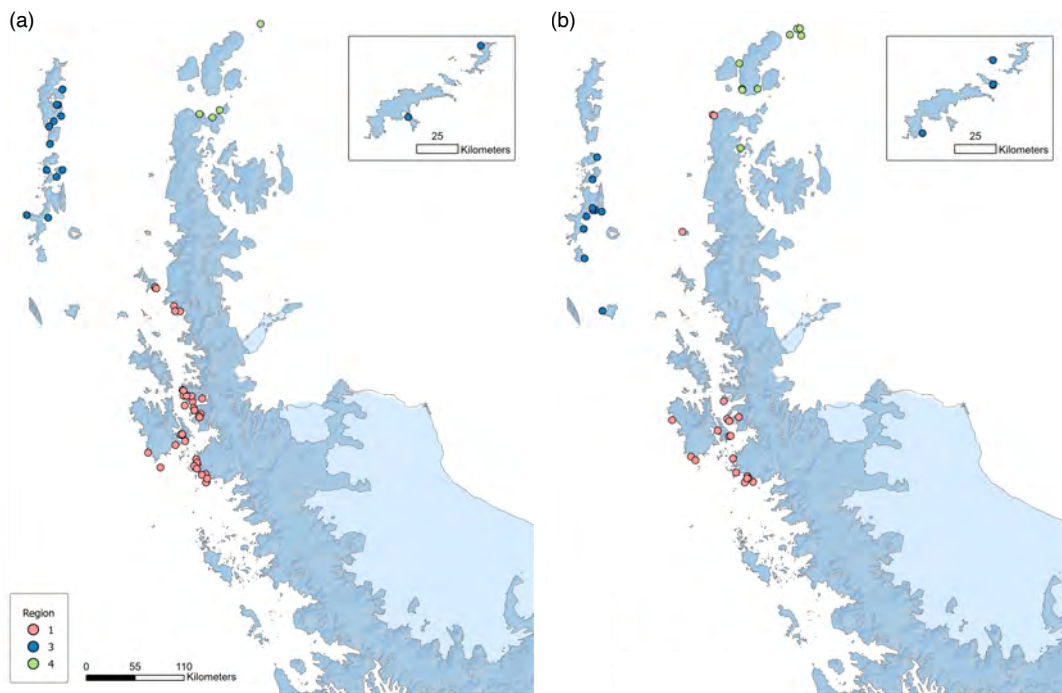
FIGURE E.3: Map of the Antarctic Peninsula showing Gentoo penguin breeding populations that were (a) included and (b) excluded from the Penguindex calculations, each population is colored by region.
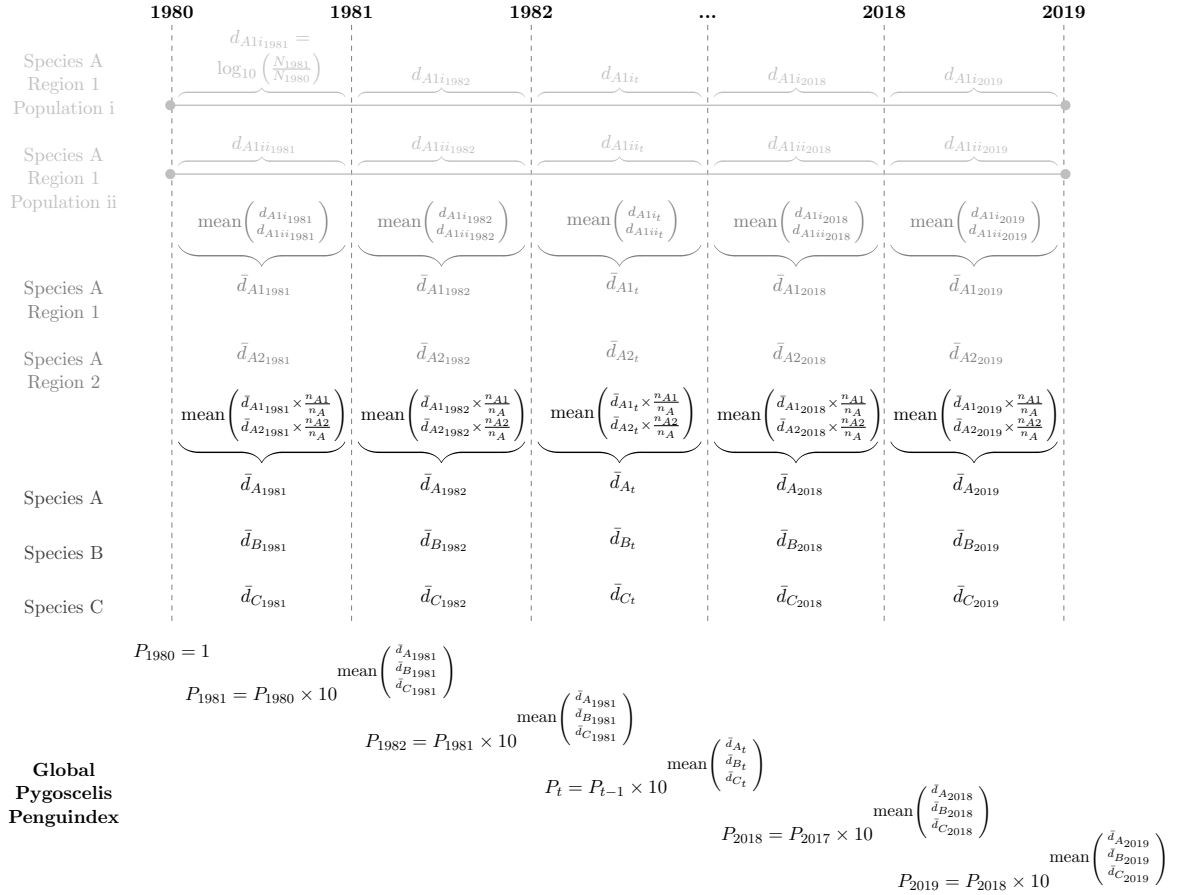
Species A Region 1 Population i

$d_{A1i_{1981}} = \log_{10}\left(\frac{N_{1981}}{N_{1980}}\right)$ $d_{A1i_{1982}}$ d_{A1i_t} $d_{A1i_{2018}}$ $d_{A1i_{2019}}$

Species A Region 1 Population ii

$d_{A1ii_{1981}}$ $d_{A1ii_{1982}}$ d_{A1ii_t} $d_{A1ii_{2018}}$ $d_{A1ii_{2019}}$

$\text{mean}\left(\begin{smallmatrix} d_{A1i_{1981}} \\ d_{A1ii_{1981}} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} d_{A1i_{1982}} \\ d_{A1ii_{1982}} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} d_{A1i_t} \\ d_{A1ii_t} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} d_{A1i_{2018}} \\ d_{A1ii_{2018}} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} d_{A1i_{2019}} \\ d_{A1ii_{2019}} \end{smallmatrix}\right)$

Species A Region 1

$\bar{d}_{A1_{1981}}$ $\bar{d}_{A1_{1982}}$ $\bar{d}_{A1_t}$ $\bar{d}_{A1_{2018}}$ $\bar{d}_{A1_{2019}}$

Species A Region 2

$\bar{d}_{A2_{1981}}$ $\bar{d}_{A2_{1982}}$ $\bar{d}_{A2_t}$ $\bar{d}_{A2_{2018}}$ $\bar{d}_{A2_{2019}}$

$\text{mean}\left(\begin{smallmatrix} \bar{d}_{A1_{1981}} \times \frac{n_{A1}}{n_A} \\ \bar{d}_{A2_{1981}} \times \frac{n_{A2}}{n_A} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} \bar{d}_{A1_{1982}} \times \frac{n_{A1}}{n_A} \\ \bar{d}_{A2_{1982}} \times \frac{n_{A2}}{n_A} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} \bar{d}_{A1_t} \times \frac{n_{A1}}{n_A} \\ \bar{d}_{A2_t} \times \frac{n_{A2}}{n_A} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} \bar{d}_{A1_{2018}} \times \frac{n_{A1}}{n_A} \\ \bar{d}_{A2_{2018}} \times \frac{n_{A2}}{n_A} \end{smallmatrix}\right)$ $\text{mean}\left(\begin{smallmatrix} \bar{d}_{A1_{2019}} \times \frac{n_{A1}}{n_A} \\ \bar{d}_{A2_{2019}} \times \frac{n_{A2}}{n_A} \end{smallmatrix}\right)$

Species A

$\bar{d}_{A_{1981}}$ $\bar{d}_{A_{1982}}$ $\bar{d}_{A_t}$ $\bar{d}_{A_{2018}}$ $\bar{d}_{A_{2019}}$

Species B

$\bar{d}_{B_{1981}}$ $\bar{d}_{B_{1982}}$ $\bar{d}_{B_t}$ $\bar{d}_{B_{2018}}$ $\bar{d}_{B_{2019}}$

Species C

$\bar{d}_{C_{1981}}$ $\bar{d}_{C_{1982}}$ $\bar{d}_{C_t}$ $\bar{d}_{C_{2018}}$ $\bar{d}_{C_{2019}}$

Global Pygoscelis Penguindex

$P_{1980} = 1$

$P_{1981} = P_{1980} \times 10^{\text{mean}\left(\begin{smallmatrix} \bar{d}_{A_{1981}} \\ \bar{d}_{B_{1981}} \\ \bar{d}_{C_{1981}} \end{smallmatrix}\right)}$

$P_{1982} = P_{1981} \times 10^{\text{mean}\left(\begin{smallmatrix} \bar{d}_{A_{1981}} \\ \bar{d}_{B_{1981}} \\ \bar{d}_{C_{1981}} \end{smallmatrix}\right)}$

$P_t = P_{t-1} \times 10^{\text{mean}\left(\begin{smallmatrix} \bar{d}_{A_t} \\ \bar{d}_{B_t} \\ \bar{d}_{C_t} \end{smallmatrix}\right)}$

$P_{2018} = P_{2017} \times 10^{\text{mean}\left(\begin{smallmatrix} \bar{d}_{A_{2018}} \\ \bar{d}_{B_{2018}} \\ \bar{d}_{C_{2018}} \end{smallmatrix}\right)}$

$P_{2019} = P_{2018} \times 10^{\text{mean}\left(\begin{smallmatrix} \bar{d}_{A_{2019}} \\ \bar{d}_{B_{2019}} \\ \bar{d}_{C_{2019}} \end{smallmatrix}\right)}$

FIGURE E.4: Schematic example of the calculation of the Penguindex. First, for each species, annual rates of change (d_t) are averaged between populations in the same region. Next, annual rates of change are averaged between regions for the same species. Lastly, annual rates of change are averaged between species. Indices are calculated from these averaged annual rates of change similarly to the LPI framework [47, 159]. Adapted from Extended Data Fig 2 in [174].

E.1.3   Quantifying Gentoo growth

To compare the overwhelming growth of Gentoo penguin populations over the last 40 years to other species undergoing population growth globally, we calculated the average percent change of each species in the public database of the most recent Living Planet Report [139] with at least one increasing population time series. Increasing populations are defined as those in which the most recent population estimate is not the minimum estimate in the time series. Percentage change of each population was calculated from the year of the minimum population size for the time series until the most recent population estimate as in [125]. These were then averaged for each increasing population time series to obtain a single percent change for each species in the database with increasing an population, amounting to 3894 species. We then obtain an empirical cumulative distribution function (ECDF) from this sample of percentage changes. The same process for calculating an average percentage change was also completed for the Gentoo penguin observations from the population time series used to calculate the Penguindex (49 increasing population time series, each with at least two observations between 1980-2019). We then calculate the percentile of the Gentoo percentage change from the ECDF obtained from the previous step. Results of the analysis of Living Planet Database species are included, with average percentage change along with the number of population time series used and the start and end years for each species are available via Zenodo: `https://doi.org/10.5281/zenodo.7573439`.

E.1.4   Population time series for inclusion in the LPI

We have provided all MAPPPD pygoscelid population observations in the format required for integration into the LPI [159]. We have unified the location names to be consistent with MAPPPD and have eliminated duplicate time series for some locations. In contrast to other species in the LPI that may have multiple time series for each location, this database contains exactly one time series for each colony in Antarctica, which represents the sum total of all publicly available data on a colony through time. Though the details of data collection may vary (e.g., direct ground counts vs. counts from UAV photographs, the composition of the survey team, etc.), all time series in the database represent the number of breeding pairs through time. These LPI-ready data are available via Zenodo: `https://doi.org/10.5281/zenodo.7573439`.

E.2 Supplementary Results

Global *Pygoscelis* Penguindex	1980	1985.8	1996.1	2015	2019		
Global Adélie Penguindex	1980	1986.5	1990.3	2005	2009.8	2019	
Global Chinstrap Penguindex	1980	1983.3	1985.2	1991	1995.7	2012.8	2014	2019
Global Gentoo Penguindex	1980	2001	2014.7	2019			
Adélie Penguindex Region 1	1980	1985.1	2006.1	2014.9	2019		
Adélie Penguindex Region 2	1980	1982.5	1987.2	1999	2003.1	2019	
Adélie Penguindex Region 3	1980	1986.3	1991.6	1994.7	2004	2019	
Adélie Penguindex Region 4	1980	1997.5	2019				
Adélie Penguindex Region 5	1980	1986.7	1990	2019			
Adélie Penguindex Region 6	1980	1988.7	2006.3	2019			
Adélie Penguindex Region 7	1980	1996	2004	2010.5	2012.1	2014.2	2019
Adélie Penguindex Region 8	1980	1988	1994.7	2004	2009.3	2019	
Chinstrap Penguindex Region 1	1980	1985	1991.4	2015.2	2019		
Chinstrap Penguindex Region 2	1980	1983.5	1989	1996.2	2010.2	2019	
Chinstrap Penguindex Region 3	1980	1983	1985.9	1990.5	1996.4	2009	2019
Chinstrap Penguindex Region 5	1980	1985.7	1998.1	2012.7	2014.1	2019	
Gentoo Penguindex Region 1	1980	2000.9	2015	2019			
Gentoo Penguindex Region 3	1980	2006.5	2019				
Gentoo Penguindex Region 4	1980	2000	2005.4	2019			

TABLE E.1: Identified eras of change (defined by change points between 1980-2019) for global *Pygoscelis*, species-level, and region-level indices.
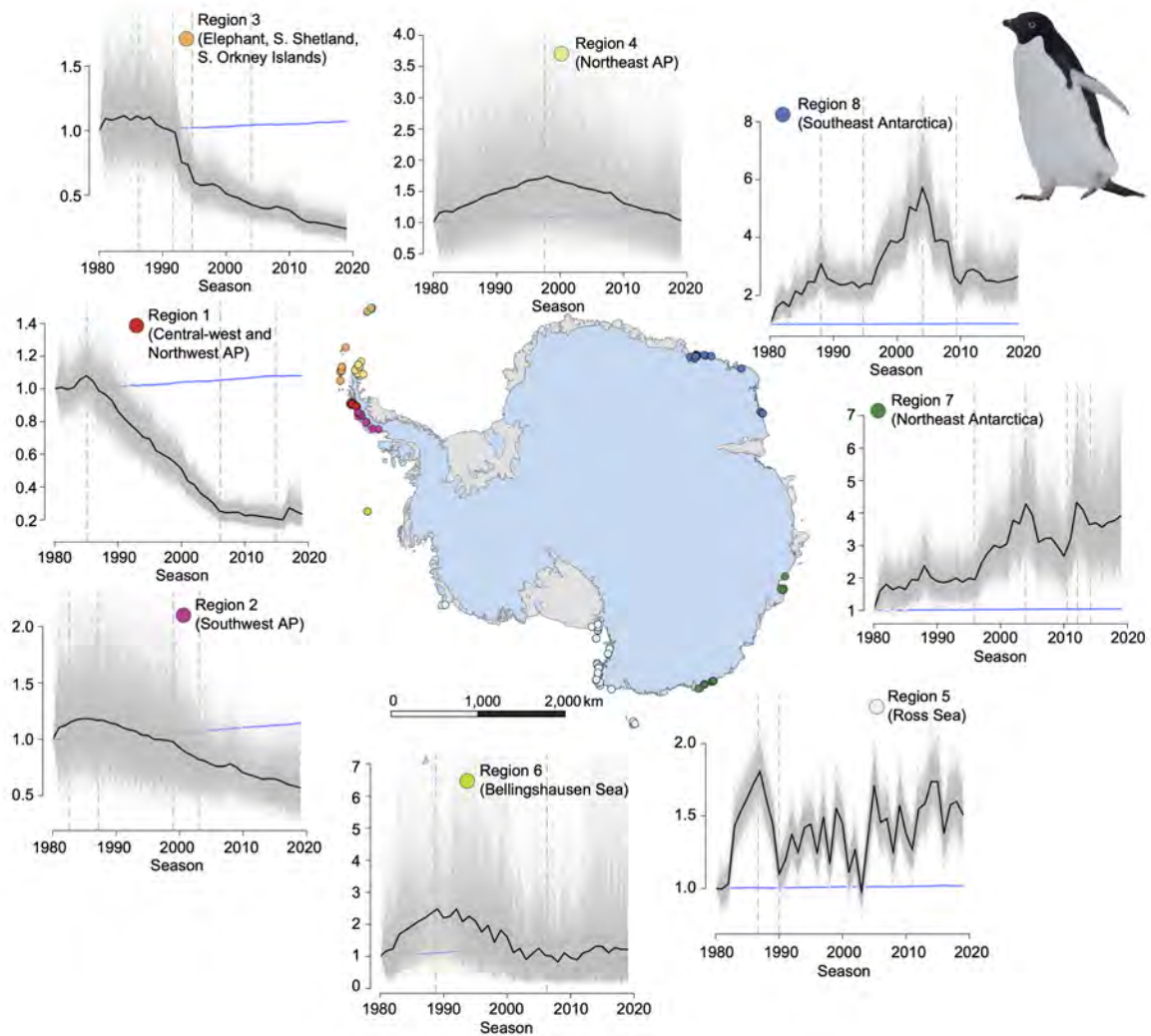
FIGURE E.5: Region-level Penguindex for Adélie penguins shown on a map of the Antarctic, including all breeding population sites used for index calculation. For each index plot, the black line denotes the mean, the white lines the 95% credible intervals, and the gray lines each iteration. Each blue line denotes the null model index.
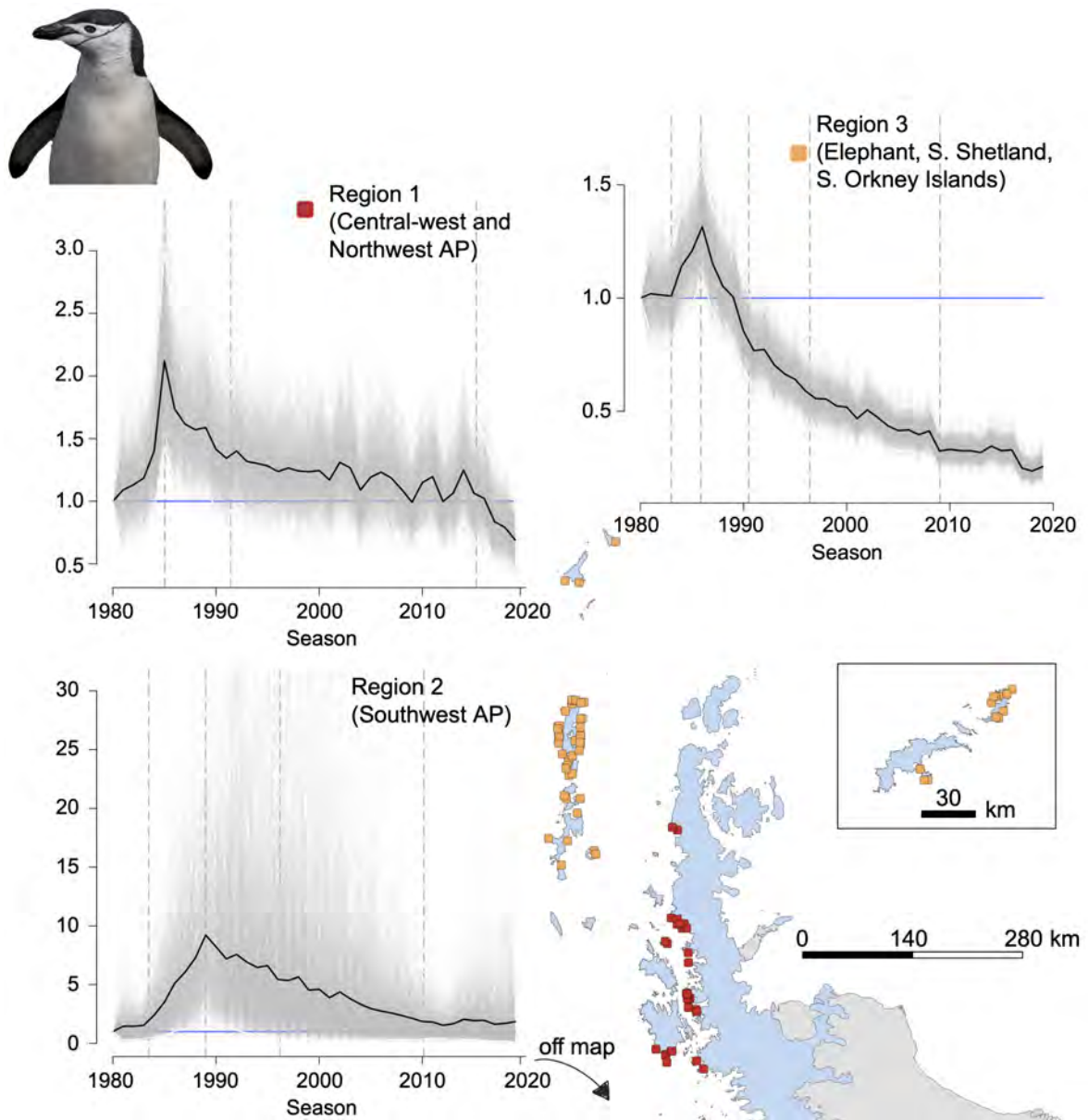
FIGURE E.6: Region-level Penguindex for Chinstrap penguins shown on a map of the Antarctic Peninsula, including breeding population sites used for index calculation. One breeding site in Region 2 (Peter I Island, Southwest AP) is not pictured on the map. Excluded is the index plot for the two Chinstrap breeding populations in Region 5 (Ross Sea). For each index plot, the 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.
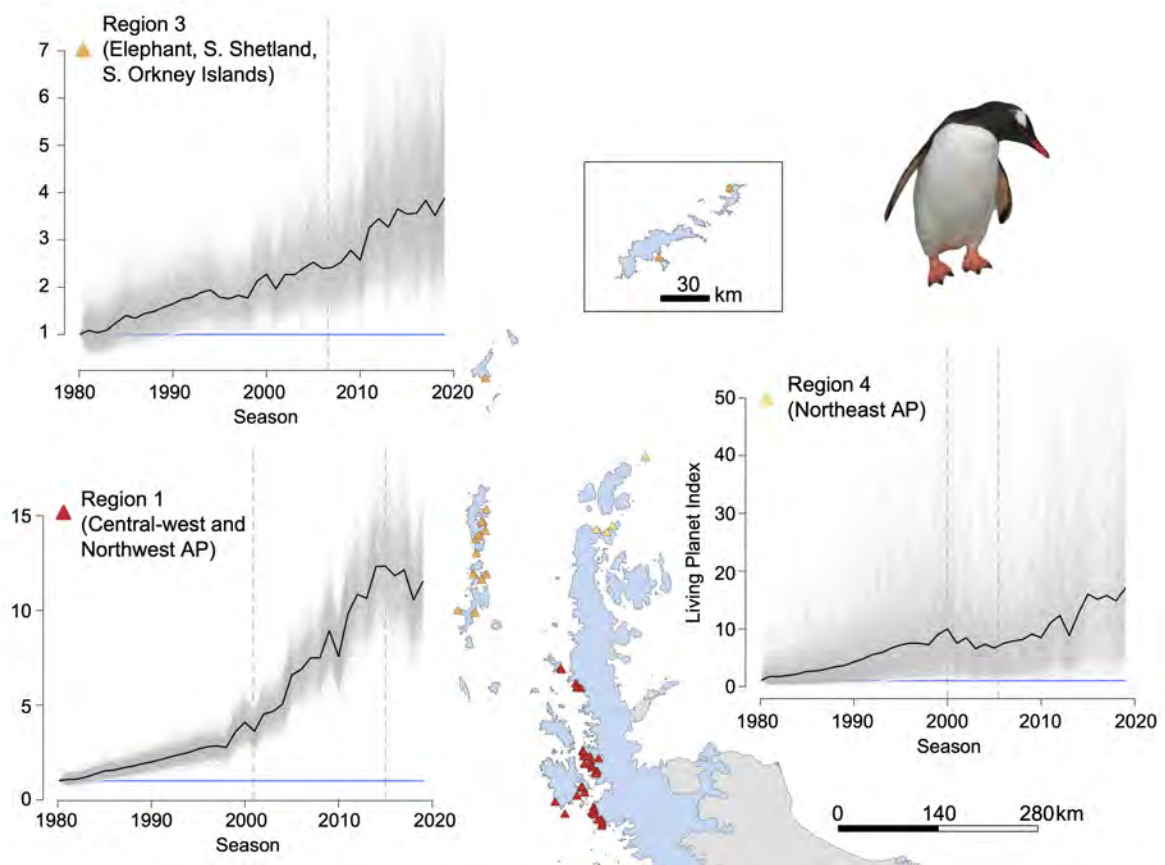
FIGURE E.7: Region-level Penguindex for Gentoo penguins shown on a map of the Antarctic Peninsula, including all breeding population sites used for index calculation. For each index plot, the black line denotes the mean, the white lines the 95% credible intervals, and the gray lines each iteration. Each blue line denotes the null model index.