

Yellowstone National Park Bison Simulation Examples

Shawna Zimmerman

20 October 2023

Change in observed allele frequency and genetic diversity

We compared three genetic data sets for change over time: 1) published allele frequencies from the Halbert et al. (2012) publication (collected 1997-2003; 46 loci), 2) genotypes for quarantined animals (collected 2018-2020; 42 loci), 3) genotypes for breeding season animals (collected in 2022; 42 loci). These data are referred to as the 2003, 2020, and 2022, respectively, from here on.

We first looked at change in allele frequencies over time (Fig.1). There were 38 loci in common across all three data sets that we could compare. We found no significant change in allele frequencies from 2003 to 2022 (Table 1).

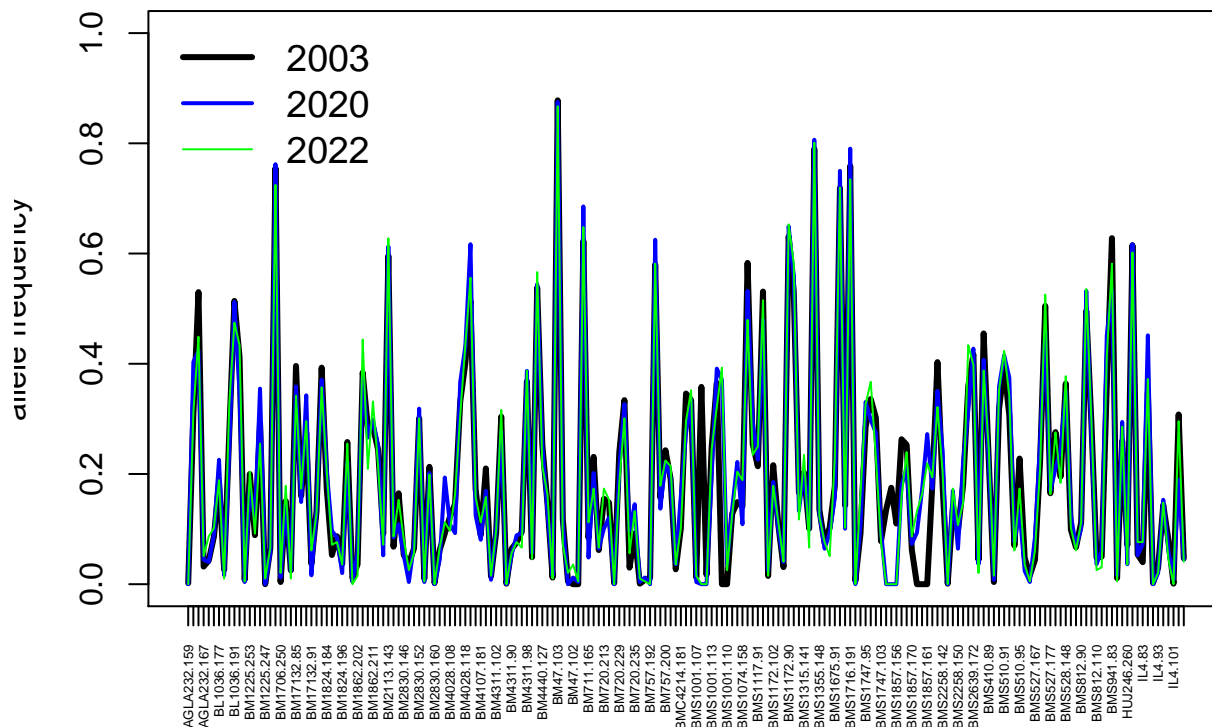


Figure 1: Line plots for empirical allele frequencies for 38 microsatellite loci from 2003-2022 for YNP bison.

There were large shifts in the the frequencies of alleles at two loci (BMS1001, BMS1857), which may reflect differences in allele calling techniques (Fig. 2). Given the step-wise mutation of microsatellite loci, it is likely that BMS1001 alleles 109 and 111 from the 2003 data are the same alleles as 108 and 110 in the 2020 and

2022 data. Similarly, the BMS1857 alleles 142, 150, and 156 in the 2003 data are likely the same as the 143, 151, and 161 alleles in the 2020 and 2022 data. These two loci were excluded from the tests for change in genetic variation over time.

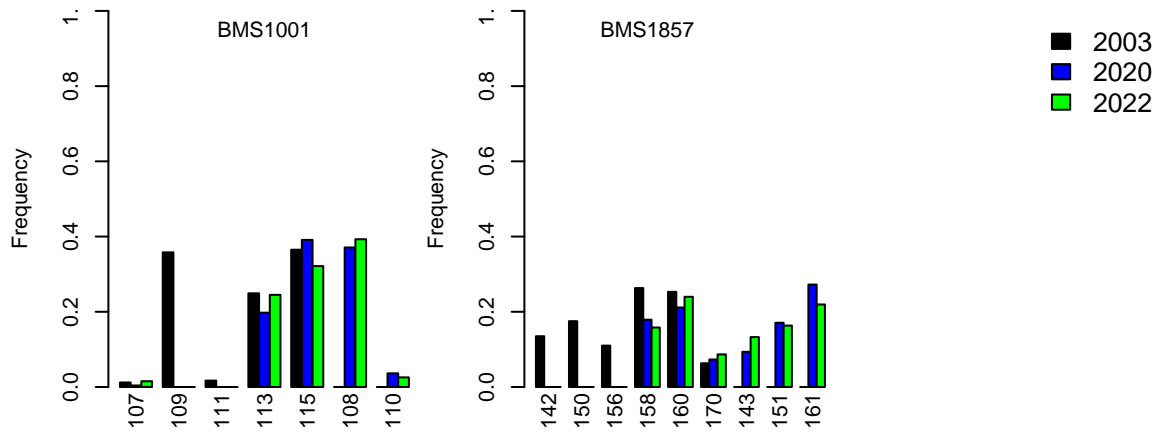


Figure 2: Barplots for empirical allele frequencies for shifted microsatellite loci allele calls from 2003-2022 for YNP bison.

Table 1: Paired t-test for change in allele frequencies for 36 microsatellite loci.

Comparison	p-value
2003 to 2020	0.99
2003 to 2022	0.98
2020 to 2022	1

We also compared expected heterozygosity values over time and similarly found no significant change from 2003 to 2022 (Fig. 3, Table 2).

Given previous findings of population structure within YNP, we also compared allele frequency (Fig. 4) and expected heterozygosity (Fig. 5) among the northern and central breeding herds for the 2003 and 2022 data and for change in those herds over time. We found no significant difference in either metric among breeding herds within a year or for the same herd across time (Table 3).

Table 2: Paired t-test for change in expected heterozygosity based on 36 microsatellite loci.

Comparison	p-value
2003 to 2020	0.19
2003 to 2022	0.91
2020 to 2022	0.17

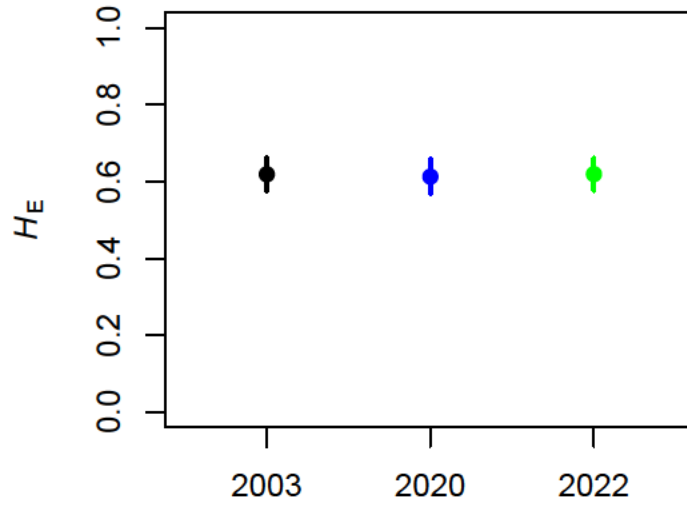


Figure 3: Empirical Expected heterozygosity values from 2003-2022 for YNP bison.

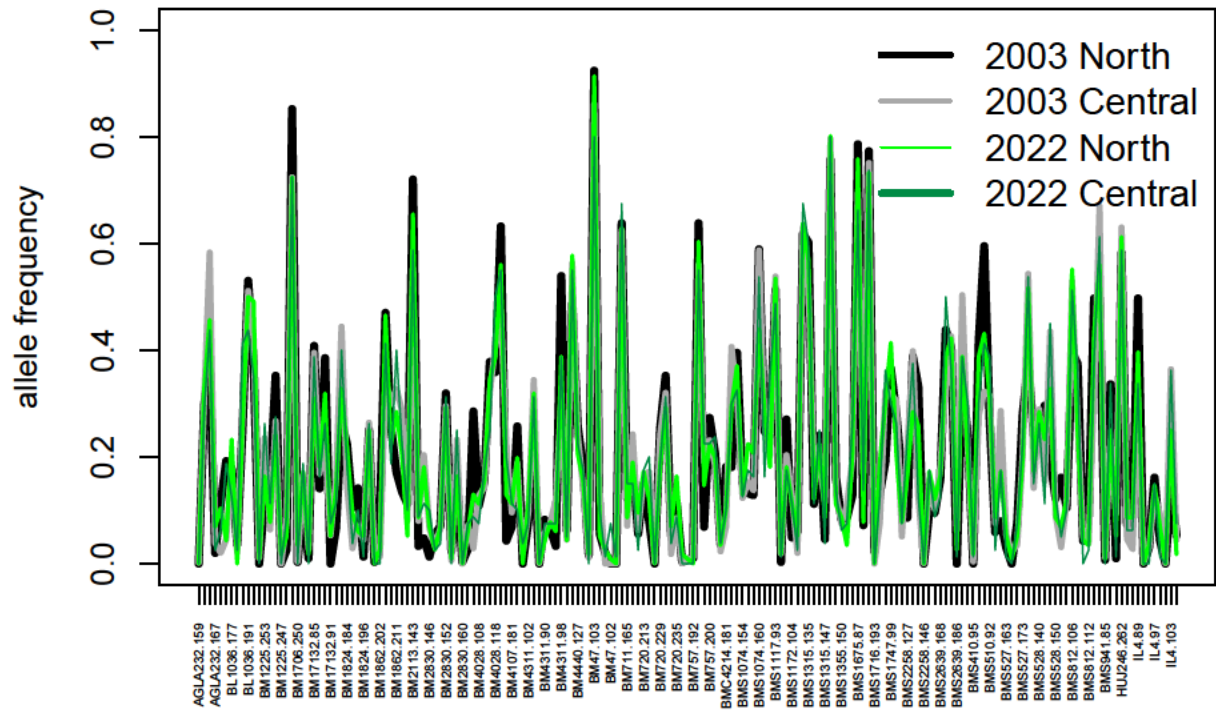


Figure 4: Empirical allele frequency values for the Northern and Central YNP bison breeding herds for 2003 and 2022.

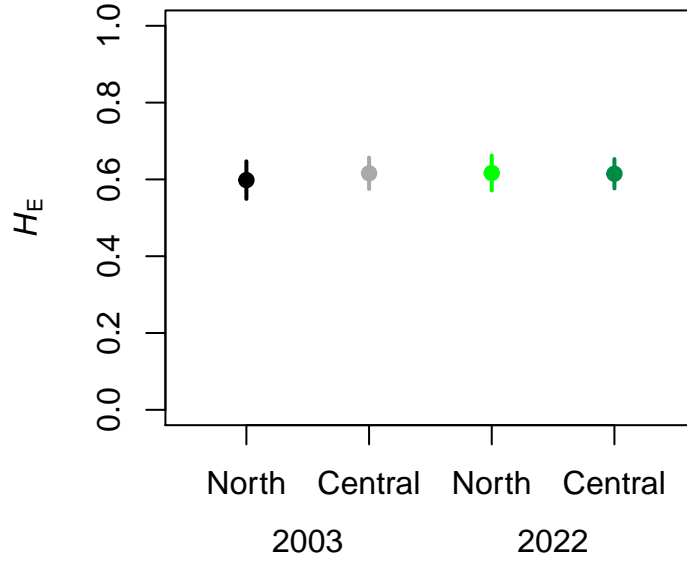


Figure 5: Empirical Expected heterozygosity values for the Northern and Central YNP bison breeding herds for 2003 and 2022.

Table 3: Paired t-test for difference in expected heterozygosity and allele frequency between Northern and Central bison herds and across years (2003 to 2022) based on 36 microsatellite loci.

Comparison	Het. p-value	Afreq. p-value
2003 North-Central	0.585	0.998
2022 North-Central	0.948	1
2003-2022 North	0.589	0.999
2003-2022 Central	0.983	0.999

Simulation model description and validation

Our simulation model is individual-based and simulates annual population change due to genetic drift through steps for reproduction, recruitment, annual survival, potential stochastic mortality, and aging (Fig. 6). The annual simulation of population change is accomplished with the `pop_sim` function. The `impact_sim` function incorporates options for simulating different population growth models, and translocation and the removal scenarios. Population growth can occur according to one of three models: exponential, logistic, or annual variation. Translocation or removals can occur after the age increment step for specified years (see below for details).

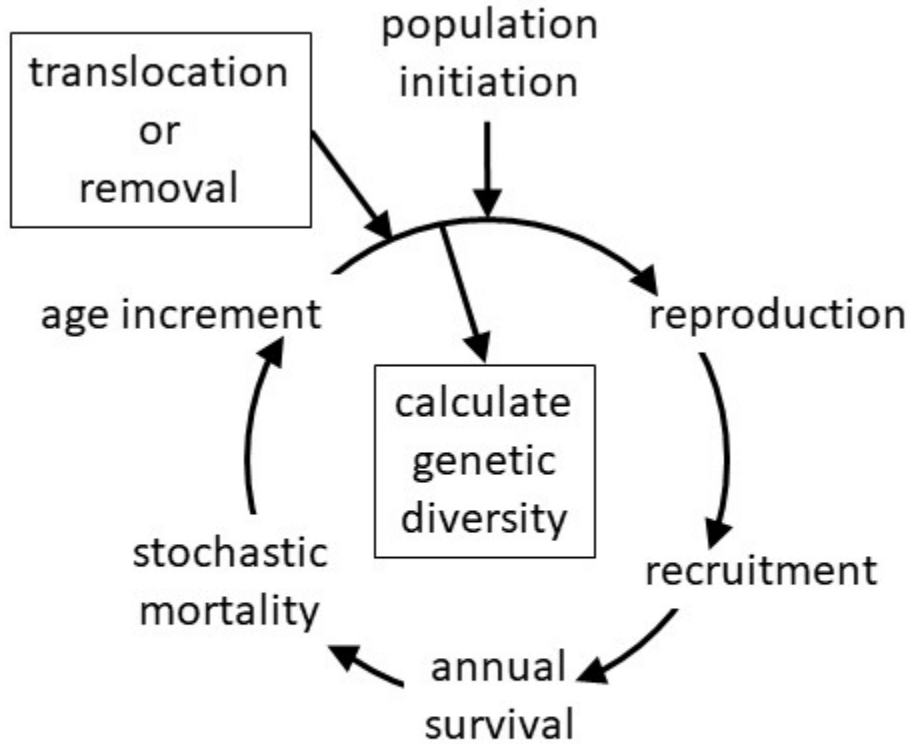


Figure 6: Annual simulation cycle.

Population initiation The simulation is initiated by simulating genotypes from allele frequency distributions, either empirical or simulated. Genotypes are simulated based on random draws from a multinomial probability distribution using allele frequency as the parameter and assuming size = 2 (assuming diploidy). Sex and age-class can be assigned directly if initial population composition is known. Alternatively, sex can be randomly assigned using a sex-ratio (defined as proportion of females in the population) and age-class can be simulated from a negative binomial or multinomial probability distribution (parameters are modifiable). The residency status of individuals is also defined for flexibility in scenarios including translocation (see below).

Reproduction Individuals are randomly selected for the breeding population based on sex and age-class specific reproduction rates as the probability of reproduction in a Bernoulli trial. Male and female individuals

selected for the breeding population are used to calculate male and female allele frequency distributions, which are used to simulate the offspring. Male and female reproduction rates are defined separately in the simulation model. The number of age-classes for each sex is dependent upon the length of the vector supplied to the model (i.e., one parameter value for each age-class). Annual variation in reproduction rates can be induced with random draws from a normal distribution by supplying an estimate of variation around the supplied means.

Recruitment Offspring are simulated from the male and female breeding population allele frequency distributions; a single allele for each locus is simulated from a multinomial distribution based on the allele frequency, assuming Mendelian inheritance. All offspring are assigned age-class zero. Sex is assigned randomly proportionate to the defined sex ratio (proportion of females). The sex ratio used to initiate the population may represent differences in survival rates while the sex ratio used in the simulation may represent differences in sex-specific birth rates. Thus, it may be appropriate to use different sex ratios initiate (pop_init) than to perform the simulation (pop_sim or impact_sim). The number of offspring recruited into the population is determined by the difference between the population size in the next generation and the population size after the annual survival step. The size of the population for each generation of simulation can be predefined by supplying a vector of integers or it can be approximated by supplying an initial population size and specifying the desired growth model (growth_model) and percentage of variation for each year (nvar; as proportion).

Annual survival Individual annual survival is determined based on sex and age-class specific survival rates used as the parameter in a Bernoulli trial. Individuals that do not ‘win’ their Bernoulli trial are removed from the population. Annual variation in survival rates can be induced with random draws from a normal distribution by supplying an estimate of variation around the supplied means.

Stochastic mortality If the population size in the next generation is smaller than the population size after the annual survival step, additional random mortality events are simulated until the N_{t+1} is smaller than N_t after the annual survival step. This assumes there are no years without recruitment and that any stochastic event will be equally applied to all sex and age-classes.

Age increment The age of each individual is incremented by 1 after all mortality events and prior to the next reproduction step, and prior to any translocation or removal interventions.

Removal Individuals can be removed from the population at any time-step in the simulation by creating vectors with the population trajectory, the years where removals are to be done, and the number of individuals to remove from the population (make_removal) and supplying these values to impact_sim. There are currently four options for how to select individuals from a source population for translocation: 1 = randomly, 2 = randomly but stratified by sex and age-class, 3 = based on relatedness (prioritizing minimally related individuals using the proportion of shared alleles), or 4 = based on relatedness but stratified by sex and age-class. Utilizing the options based on relatedness increase computation time, particularly for large populations.

Calculate genetic diversity The model calculates and stores four measures of genetic diversity for each time-step: allelic richness ($A_{\sim R}$), mean number of alleles (A), observed heterozygosity (H_O), and expected heterozygosity (H_E). Estimates of allelic richness are highly sensitive to sample size and so are rarefied to a defined minimum sampled size (default=10). The population size (N), effective population size (N_E), the number of breeders (N_B), the number of mortalities (N_d), the number of males (N_m), and the number of females (N_f) are also stored for each time-step.

Validation We validated our simulation model by demonstrating that heterozygosity is lost over time as theoretically predicted for a stable (Fig. 7a) and fluctuating (Fig. 7b) population size assuming all other ideal population characteristics (discrete generations, equal reproduction, even sex ratio). We further show that when we relax these assumptions the model behaves as expected: if we allow overlapping generations (Fig. 7c,d) the loss of genetic diversity is slower than the theoretical ideal population and if only 50% of individuals reproduce each year the loss is more rapid than the theoretical ideal population (Fig. 7e,f).

We also include three options for population growth models: logistic (Fig. 8a), exponential (Fig. 8b), and annual variation (Fig. 8c).

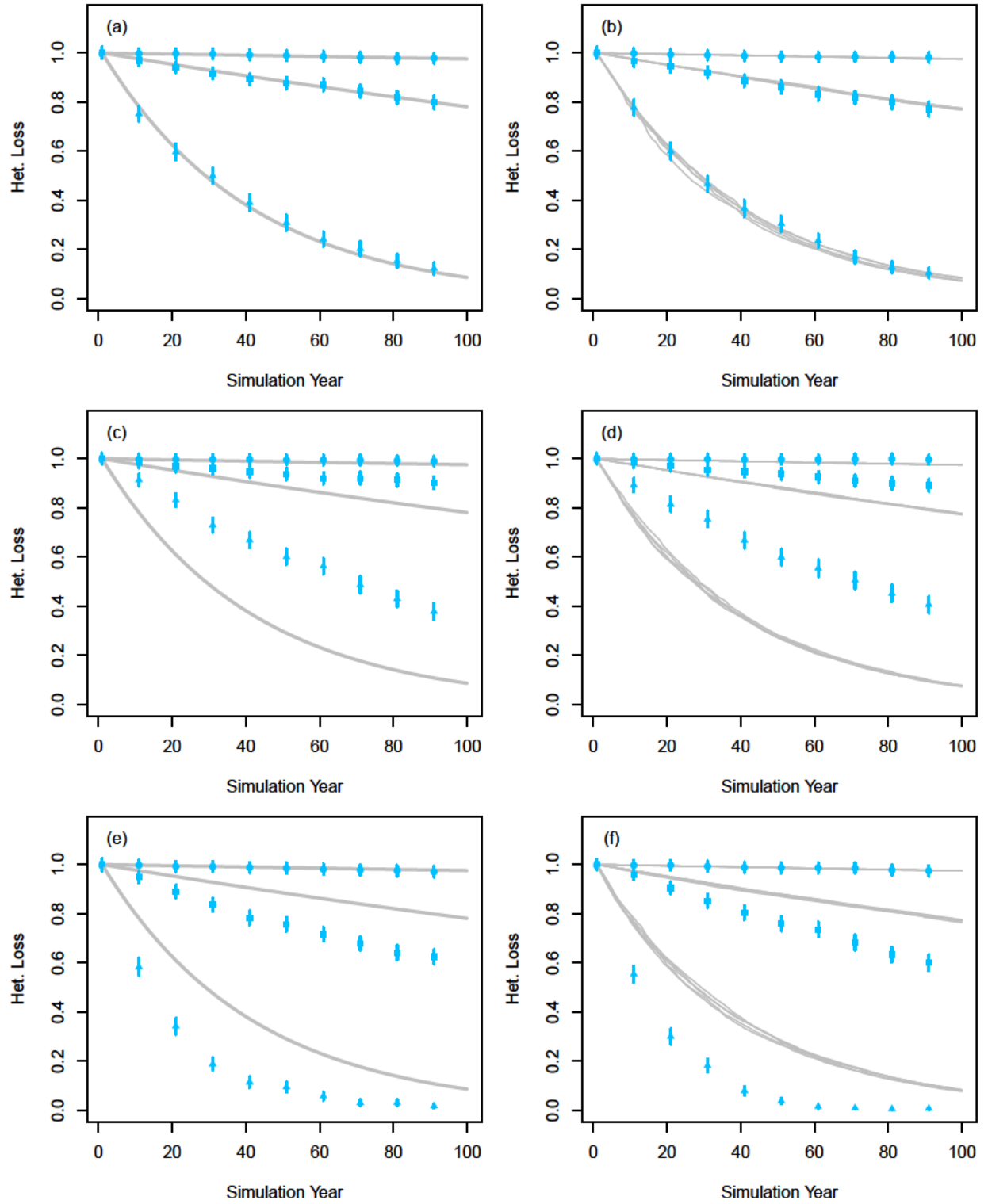


Figure 7: Theoretical genetic diversity loss (gray line) assuming discrete generations, equal reproduction, even sex ratio under a stable population (a) and fluctuation population (b) compared to predicted loss in our simulation model (points). Predicted genetic diversity loss is slowed relative to theoretical when generations overlap (c,d) and increased when only some individuals reproduce each year (e,f).

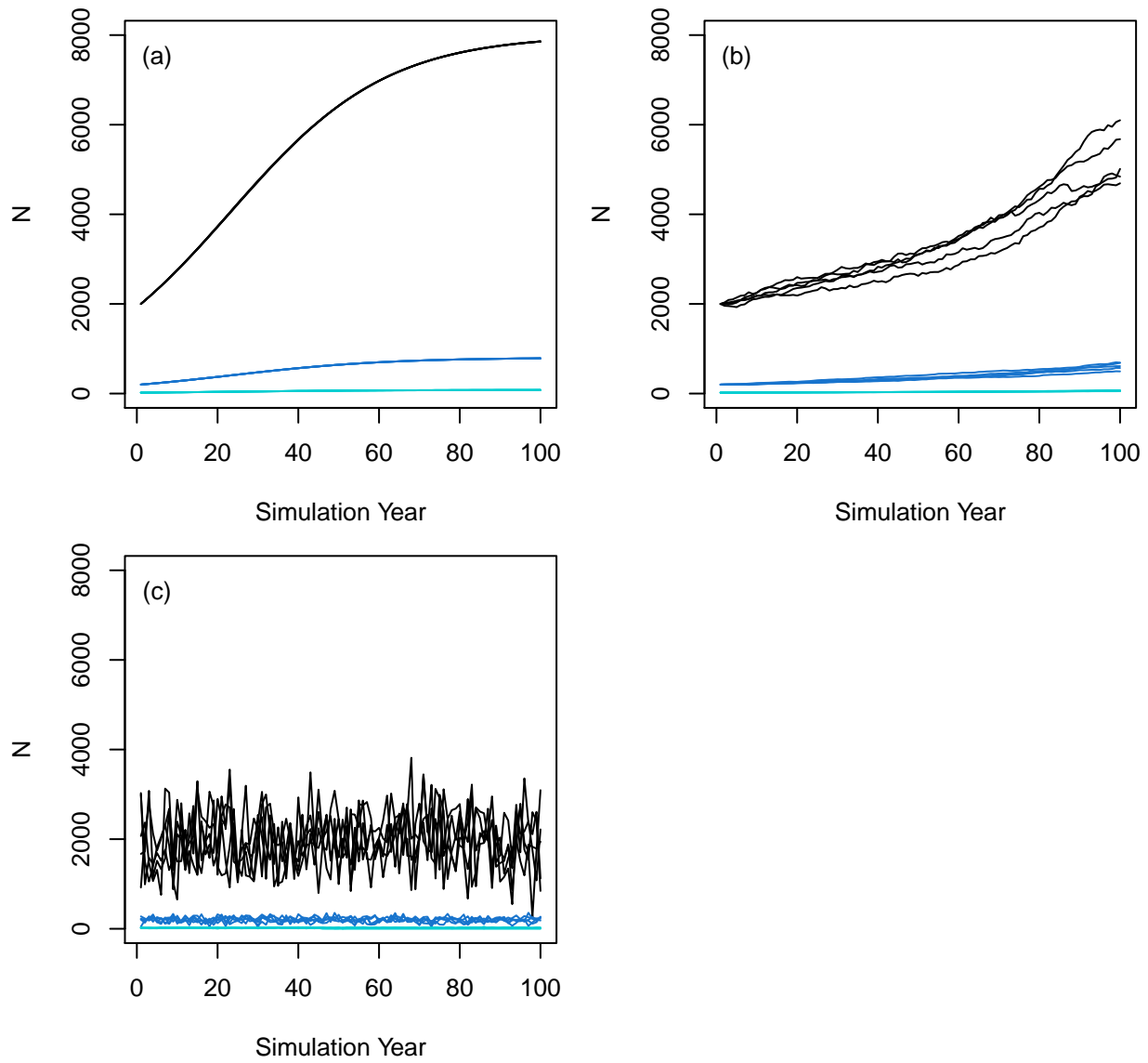


Figure 8: Population growth models: a) logistic, b) exponential, c) annual variation

Application of simulation model to YNP bison

We applied our simulation model to the YNP bison herd using the genetic data from 2022 and YNP specific demographic parameters. We used the mean and standard deviation of provided age-specific demographic parameters to inform our model. This included female survival and reproduction rates, population growth rate, and sex ratio. Male survival was assumed the same as female survival (per provided demographic information). Male mean reproduction rates were based on values reported in the PVA, while a measure of variability was simulated based on the female reproduction rate standard deviations. Annual variation is induced in the demographic parameters as random draws from a normal distribution using the empirical means and standard deviations as parameters. Values <0 or >1 are illogical and corrected to 0 and 1, respectively.

The starting age distribution for the herd was based on percentages published in The Bison of Yellowstone National Park NPS Scientific Monograph No. 1: 0 = 16%, 1 = 11%, 2 = 2.5%, 3 = 5%, 4-20 = 65.5%.

For our initial simulations we assumed the initial herd size was 1500 animals and simulated 100 years of change over 5 replicates. We included three scenarios simulating predicted genetic diversity loss as examples:

1. Maintaining a stable herd size of ~1500 animals.
2. Logistic population growth with a carrying capacity of 3000 animals.
3. Logistic population growth with a carrying capacity of 3000 animals and removing 500 animals every 8 years.

1) Stable herd size

In this scenario we assumed a stable herd size of ~1500 animals with 5% annual variation. Variation in herd size across years and replicates is induced by setting the `nvar` parameter in `impact_sim` to the amount of variation desired (as proportion).

The relatively large population size and long, overlapping generations of bison result in very slow genetic diversity loss relative to a theoretical ideal population (Fig. 9). Our model predicts a 98.43% maintenance of heterozygosity and 100% maintenance of allelic diversity.

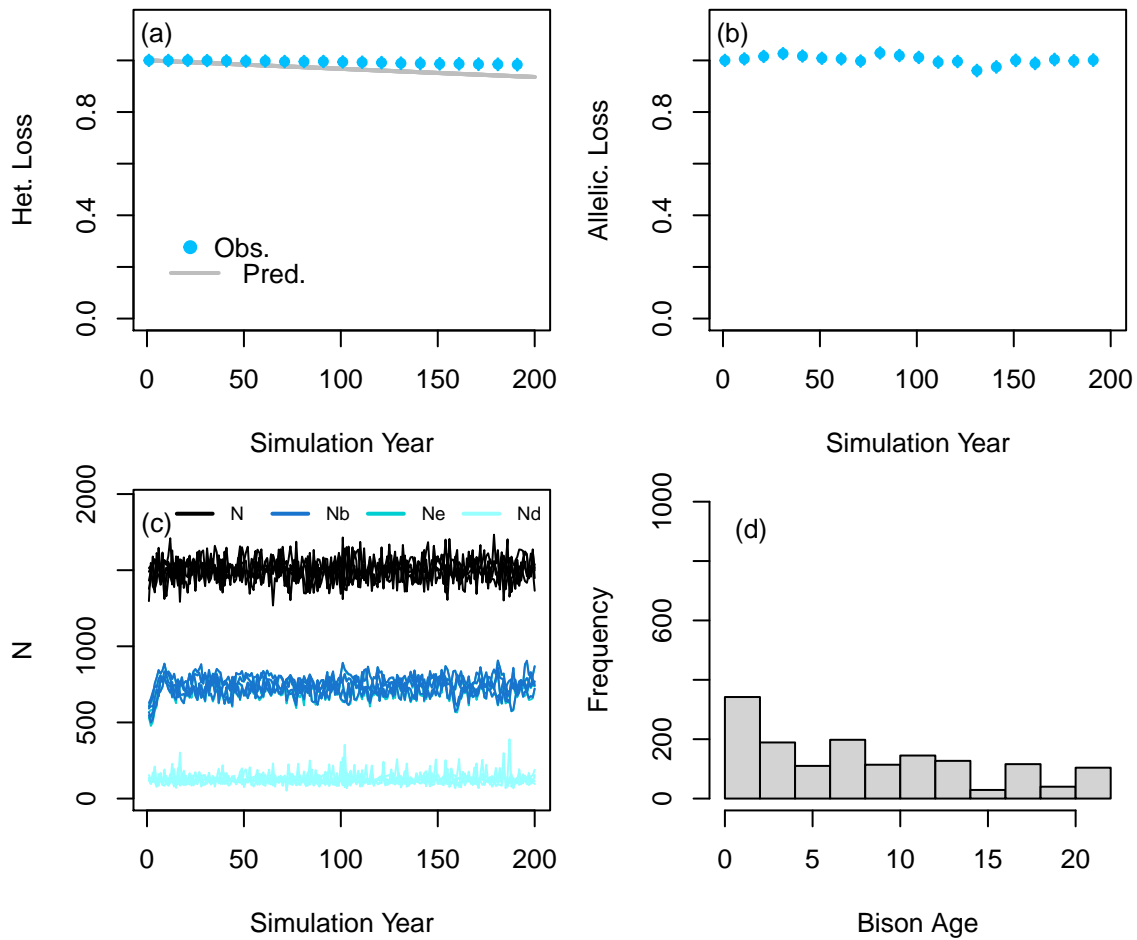


Figure 9: Stable herd size with 5% annual variation. a) predicted heterozygosity loss compared to theoretical, b) predicted allelic loss, c) population trajectory, d) final simulated age structure of the herd.

2) Logistic growth

In this scenario we assumed logistic growth from an initial herd size of ~1500 animals and a carrying capacity of 3000 animals. The logistic growth model is selected by setting the `growth_model` parameter to “logistic.” As in the stable population simulation, variation in initial herd size across years and replicates is induced by setting the `nvar` parameter in `impact_sim` to the amount of variation desired (as proportion).

As above, the relatively large population size and long, overlapping generations of bison result in very slow genetic diversity loss relative to a theoretical ideal population (Fig. 10). Our model predicts a 99.41% maintenance of heterozygosity and 99.16% maintenance of allelic diversity.

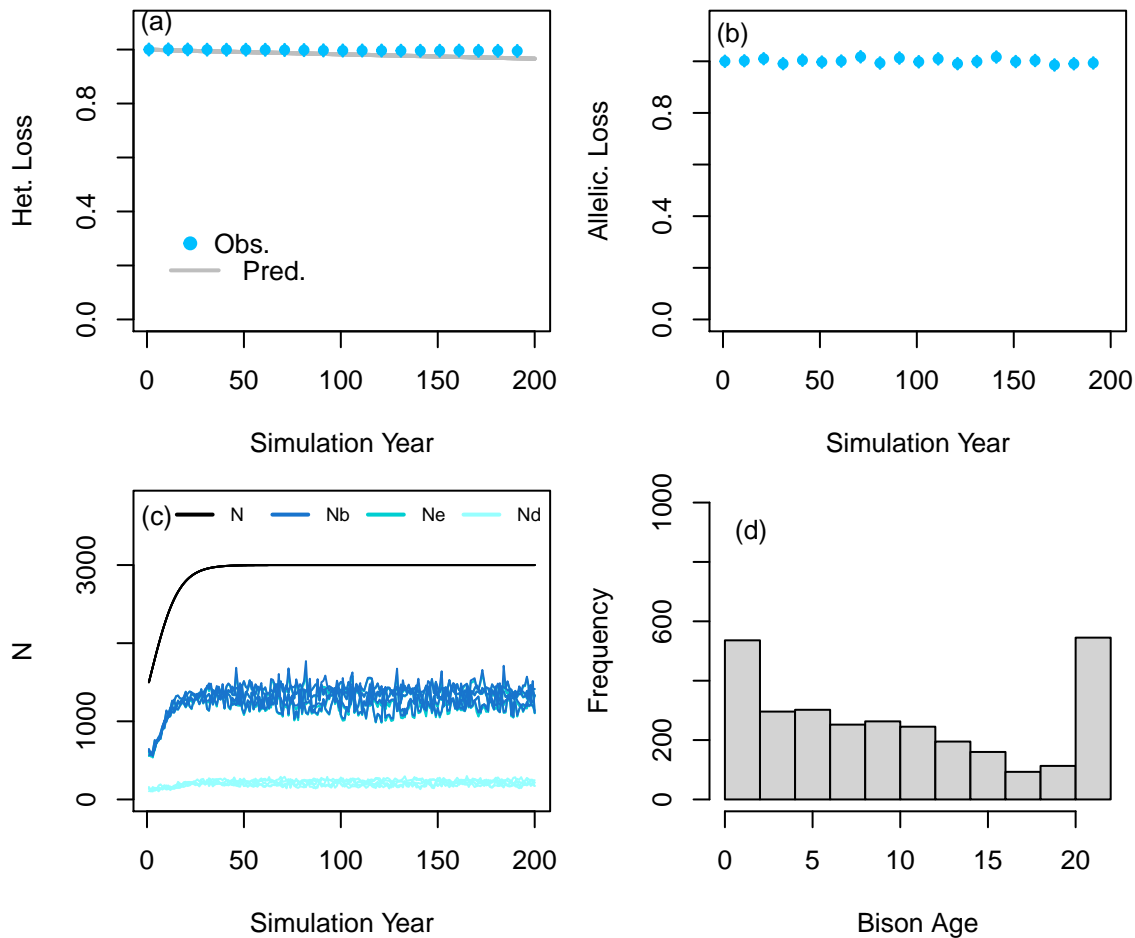


Figure 10: Logistic growth with 5% annual variation ($N_{init} = 1500$, $K = 3000$). a) predicted heterozygosity loss compared to theoretical, b) predicted allelic loss, c) population trajectory, d) final simulated age structure of the herd.

3) Logistic growth with removals

In this scenario we assumed logistic growth from an initial herd size of ~1500 animals, a carrying capacity of 3000 animals, and assumed 500 animals were removed from the herd every 8 years.

For a removal scenario, we first use the `make_removal` function to create a names list object including population trajectory under a logistic growth model that accounts for removals (`pop_size`), a vector of removal years (`years`), and the number of individuals that should be removed from the herd for each removal year (`remove`). The three lists should then be supplied to the `n`, `ryears`, and `remove` parameters in the `impact_sim` function. Importantly, `growth_model` in the `impact_sim` function must be set to “annual_variation” (not “logistic”). This will prevent simulation model from overwriting the population vector that accounts for removals. Variation across replicates and years can be still be induced using `nvar`. The composition of animals removed from the herd was based on target sex and age-class ratios. The supplied proportion of females (`propf`) is used to maintain the desired balance of sexes. To target age-classes, they must be defined (`age_classes`) and the proportion of the herd desired within each defined age class must be specified (`propt`).

As above, the relatively large population size and long, overlapping generations of bison result in very slow genetic diversity loss relative to a theoretical ideal population (Fig. 11). Our model predicts a 98.69% maintenance of heterozygosity and 97.89% maintenance of allelic diversity.

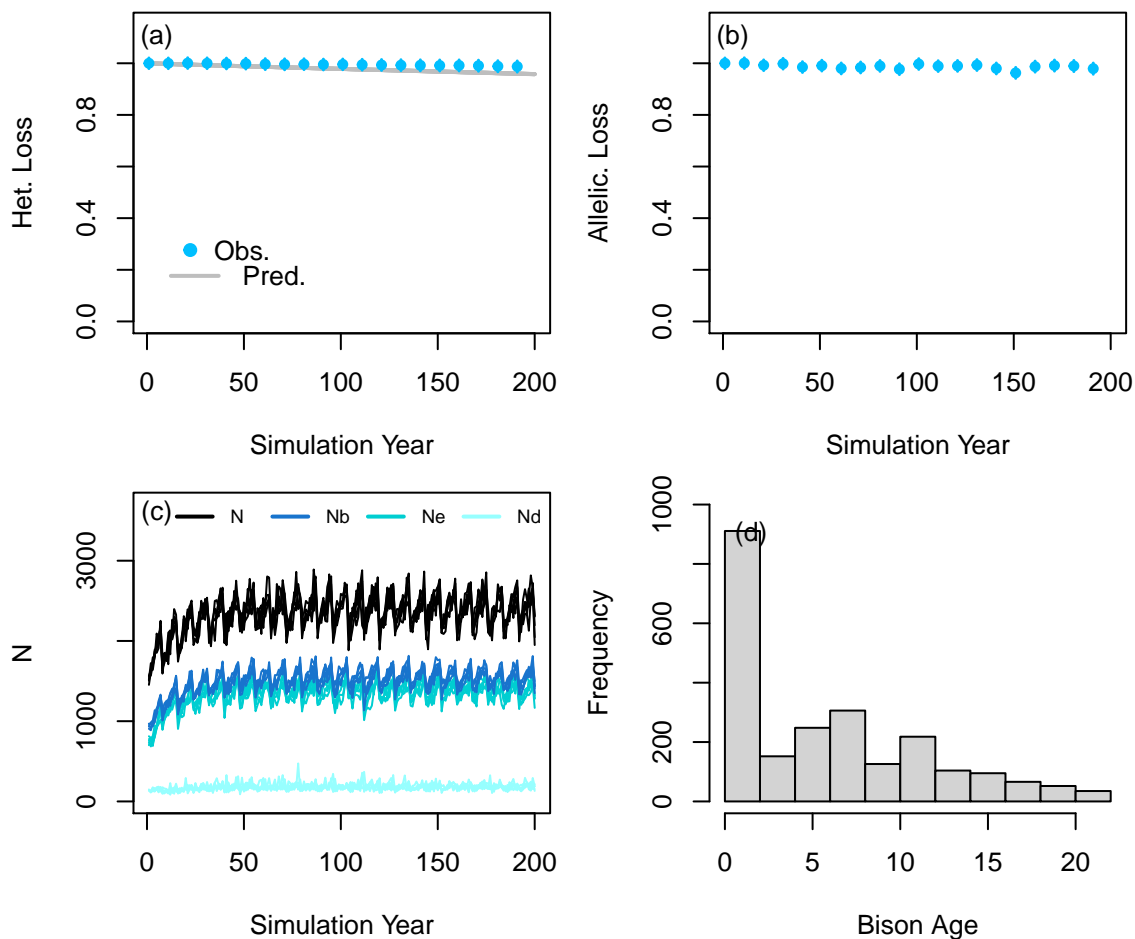


Figure 11: Logistic growth with 5% annual variation ($N_{init} = 1500$, $K = 3000$) and removing 500 animals every 8 years. a) predicted heterozygosity loss compared to theoretical, b) predicted allelic loss, c) population trajectory, d) final simulated age structure of the herd.

Simulation of genetic change from 2003-2023 given observed population trend and known removals.

We simulated the predicted change in genetic diversity assuming ~5% annual variation from observed population trend from 2003-2023 and including known removals. Our model predicts virtually no loss in genetic diversity (Fig. 11a and 11b) and tracks the observed trend and removals (Fig. 11c). Our model predicts a 99.92% maintenance of heterozygosity and 99.14% maintenance of allelic diversity. Expected heterozygosity was 0.619 in 2003 and 0.6195 in 2022, consistent with our model prediction.

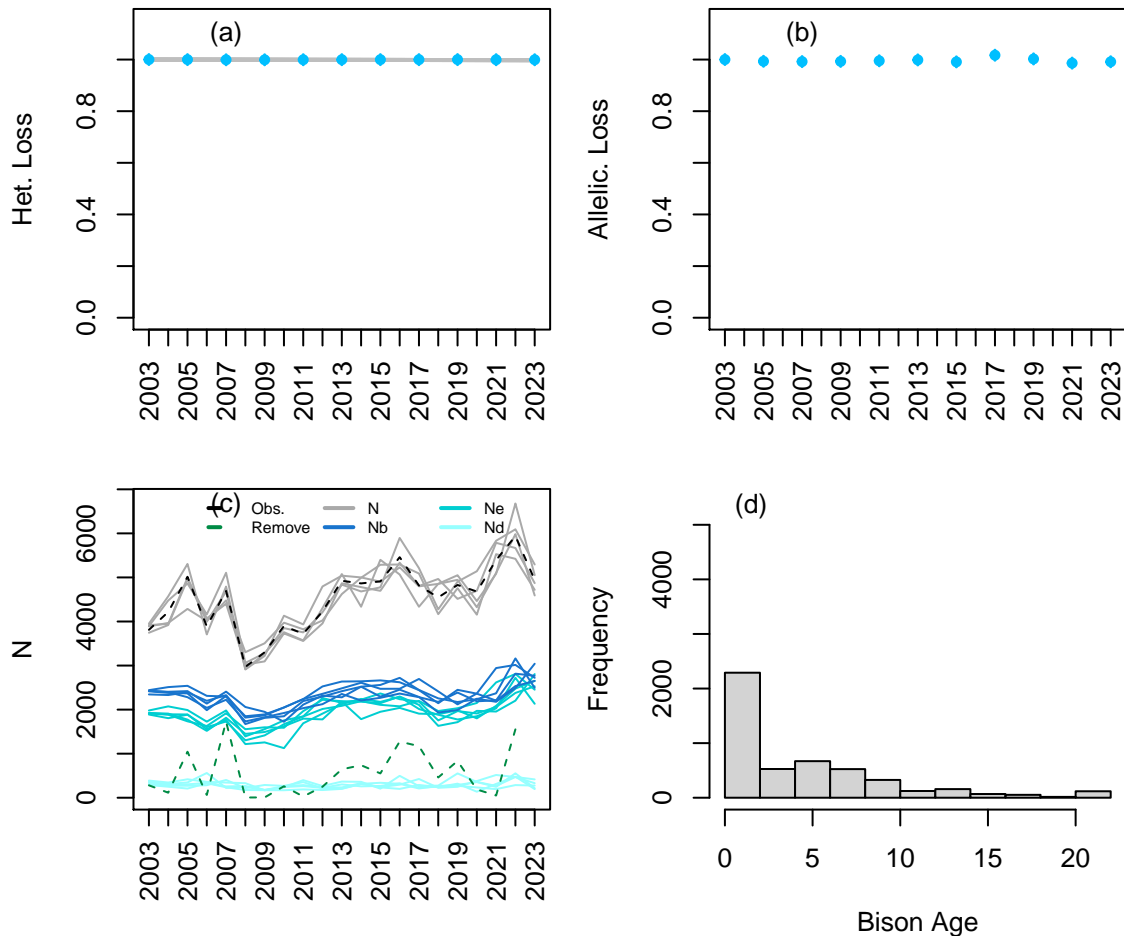


Figure 12: Simulation of predicted change in genetic diversity from 2003-2023 for observed population trend and known removals. a) predicted heterozygosity loss compared to theoretical, b) predicted allelic loss, c) population trajectory, d) final simulated age structure of the herd.

Questions:

1. Should we separate the breeding herds for simulation?
2. Should we combine data from 2019-2022? And the forthcoming 2023?
3. Which values should we use for male reproductive success? Currently using values approximated from the PVA.
4. Should we include male dominance in our simulation model?
5. Should we consider different methods to select individuals for removal?