2017

EXPLORING POPULATION CHANGE DETECTION BY MONITORING EFFECTIVE NUMBER OF BREEDERS

Brian Trethewey
University of Montana

EXPLORING POPULATION CHANGE DETECTION BY MONITORING EFFECTIVE NUMBER OF BREEDERS

By

Brian Richard Trethewey

BS. Biotechnology, Montana State University, Bozeman, Montana, 2009

Thesis

presented in partial fulfillment of the requirements
for the degree of

Masters of Science
in Computer Science
The University of Montana
Missoula, MT

December 2017

Approved by:

Scott Whittenburg, Dean of The Graduate School
Graduate School

Alden Wright, Co-Chair
Computer Science, University of Montana

Brian Hand, Co-Chair
Division of Biological Sciences, University of Montana

Jesse Johnson
Computer Science, University of Montana

Trethewey, Brian, M.S., Fall 2017 Computer Science

Abstract:

Co-Chairperson: Alden Wright

Co-Chairperson: Brian Hand

Detecting if a population is in decline is an important objective for biologists and conservationists who are monitoring threatened populations. As genetic methods improve effective population size (N_e) and effective number of breeders (N_b) continue to gain popularity as a way to monitor species. Using simulated populations and linkage disequilibrium, we explored detecting population decline through N_b in age structured populations. Through comparisons of sensitivity (1 – false negatives) and specificity (1- false positives) over 1000 replicates, we explored how factors such as starting N_b, number of SNPs, number of individuals sampled, number of breeding cycles monitored, and rate of decline affected the ability to detect changes in the population. Overall, we found N_b can be an effective metric for detecting population declines, if some care is taken during study design to avoid certain conditions. Although specificity did not vary greatly, sensitivity was much more reactive to changes in the factors tested. Under-sampling of the population (< true N_b), insufficient number of breeding cycles monitored (<7 cycles) and low levels of decline (e.g. <7%), are all detrimental to detection of population change.

Introduction

Population decline is a potential strong indicator of an extinction event (Collen et al. 2011). Small populations are especially sensitive to genetic effects such as inbreeding depression and lack of allele diversity. These effects make the populations especially susceptible to extirpation (Mace et al. 2008). Regular, usually annual, sampling of a species of interest, is used by conservationists to monitor populations and create management plans to help these populations recover and remain at safe population levels.

Genetic drift describes stochastic changes of allele frequencies in a population during reproduction, independent from selection or mutation (Hartl and Clark 1998). Assumptions such as random mating, constant population size, and equal sex ratio, allows changes in allele frequency in the next generation to be modeled as a multinomial distribution. This concept is used in the Wright-Fisher model to simulate genetic drift in populations.

The most straightforward way to detect changes in the size of a population is to take a census of the adults. The adult census size (N_c) can be also be determined through estimation techniques such as mark and recapture. These methods can be inaccurate or difficult depending on the population being studied (Luikart et al. 2010). For example, in some fish populations with large, difficult to access habitats, such as the ocean, population census size can be difficult to estimate.

Effective population size (N_e) is traditionally defined in terms of a Wright-Fisher model that has the same rate of genetic drift as the observed population (Hartl and Clark 1998). N_e can also describe the size of an Wright-Fisher population that loses heterozygosity at the same rate as

the observed population. N_e estimations have a relationship to N_c based on many factors such as the life history of the species, reproductive success, and family structure. (Cornuet and Luikart 1996; Perrier et al. 2016). N_e can be calculated from the distribution of alleles among individuals of a population. Not only is it another metric for monitoring population size but it also includes important information about the genetic diversity of a population (Perrier et al.,2016), which can indicate susceptibility to genetic drift, inbreeding depression, or loss of adaptive ability (Hare 2011; Mace et al. 2008).

Effective number of breeders (N_b) is a refinement of N_e, adding the stipulation that all sampled members belong to the same cohort (are produced during the same breeding cycle). The number and genetic composition of the reproductively successful individuals of the parental generation is the primary factor reflected in the calculation of N_b (Whiteley et al. 2017), hence N_b can be calculated from a non-breeding cohort. Because N_b data is processed the same as N_e only with a shorter timeframe, N_b is a more accurate measure of the population's current genetic heath. The cohort restriction of N_b is especially applicable for some species such as bull trout (*Salvelinus confluentus*) that can be sampled at spawning ensuring that every sampled individual is from the same cohort (Perrier et al. 2016).

N_b also provides more information on individual species' breeding habits, such as sexual selection and family structure (Whiteley et al. 2015). Whiteley et al. (2015) also found there is no simple direct relationship between N_b and N_c in their salmonid populations, suggesting that N_b is affected by much more than just adult abundance. In contrast Bernos, and Fraser (2016) found that N_b and N_c were positively correlated in wild brook trout (*Salvelinus fontinalis*) populations although the N_b/N_c ratio fluctuated over time.

Estimation techniques for both N_e and N_b suffer from bias when used on age structured populations. This bias in N_b's case can be corrected through use of the estimate of an empirically derived N_b/N_e ratio (Waples et al. 2014). For N_e this bias is not as easily or efficiently corrected as the severity depends on the number of cohorts being sampled.

Estimates of the N_e/N_c and N_b/N_e ratios can be calculated from life history traits, primarily longevity and sexual maturity. Waples et al. (2014) derived an empirical relationship between N_b, N_e and N_c, and included it in a program called AgeN_e, based on the lifetables of the species (Waples et al. 2011, Waples et al. 2013). These lifetables describe survival and reproductive rates of a species as they age, usually on a per gender basis. Although these derived methods are useful for rough estimations, they ignore the stochastic forces (e.g. genetic drift), to address the effect of these forces requires simulation over purely deterministic approaches.

Simulation modeling is an important tool in evolutionary biology. Simulations help test hypotheses under controlled and ideal situations that do noyt occur in a natural environment where there are unknown and uncontrollable factors influence the results. One class of simulator, known as forward-time simulators, are used to simulate successive reproduction in usually isolated populations (Yuan et al. 2012). Forward simulators provide insight into the evolutionary processes by modeling the stochastic forces involved, such as random mating and survival. Forward simulators become increasingly complex as additional genetic and environmental factors are added to the simulation (Peng et al. 2005).

Simupop is a highly extensible forward time simulator based on the Wright-Fisher model (Peng and Kimmel 2005; Peng and Amos 2008). Waples et al (2014) developed a population genetics model that extended Simupop to simulate multi-generational age structured populations, implemented by using lifetables to describe the population's gender and age structure.

The most used method to calculate N_b and N_e uses linkage disequilibrium (LD) (Luikart et al. 2010). The LD method calculates N_e or N_b by measuring the statistical unlikeliness of allele occurrences between pairs of loci in the population (Waples and Do 2008; Hill and Robertson 1968; Luikart et al. 2010). This is to say assuming all loci are independent, LD is based on the measuring how the observed occurrences deviate from the expected frequencies. The LD method was previously seen as less accurate than the more conventional temporal method when the number of individuals sampled was less than N_e (Luikart et al. 2010). Advances by Waples (2006) corrected this bias empirically. The LD method heavily weights the effects of genetic drift (Luikart et al. 2010). This makes the LD method better for applications requiring fast responses such as monitoring threatened species.

Sensitivity, specificity, and receiver operating characteristic (ROC) curves were used to quantify and illustrate the efficacy of using N_b to detect changes in population size, because the N_b test is a binary classifier system. In this context sensitivity is the probability that a change in population size is detected when the population size is changing, and specificity is the probability that no change in population size is detected when the population size is stable. For specificity, higher values indicate increased detection of stability in the population. Higher sensitivities indicate increased detection of changes in the population. ROC curves are used to illustrate the tradeoffs between sensitivity and specificity by graphing the changes in sensitivity as false positive rate (1 - specificity) is changed. The ROC curve provides a visualization of how well a change in population size can be detected using N_b and how close the N_b test is to a perfect test, one with sensitivity and specificity of one.

Hypotheses

N_b can be an effective metric for monitoring changes in population size. We want to determine under what circumstances it can be applied to successfully monitor a population, and then evaluate the effect of some factors on the sensitivity and specificity of using N_b to detect decline in N_c. This was accomplished by evaluating the following hypotheses, with the assumption of no mutation or selection, and no environmental factors.

Box 1: Factors explored in this study

Factors:

- Number of SNPs evaluated
- Starting N_b
- Number of breeding cycles monitored
- Severity of decline
- Number of individuals sampled

Null 1: Changes in census size (N_c) will be undetectable using N_b over a short timeframe (≤10 breeding cycles) in small (≤100N_b), near Wright-Fisher populations.

Alternative 1: Changes in N_c are detectable using N_b over a short time span in small, near Wright-Fisher populations.

Null 2: Stability of N_c is undetectable over a short time scale (≤10 breeding cycles) in small (≤100N_b), near Wright-Fisher populations.

Alternative 2: Stability of N_c is detectable over a short time span in small, near Wright-Fisher populations.

We designed our hypotheses to be two sided because depending on the study either growth or depression may be of interest.

We also tested the robustness of these alternatives by weakening several of the assumptions of the Wright-Fisher model and how some common factors affect detection rates (Box 1).

The first null hypothesis is tested by simulating a stable population to burn-in, then enforcing a 5%, 7%, 10%, and 15% decline from that point forth. Each simulated decline is sampled at each breeding cycle and N_b is calculated from the youngest cohort for a span of five, seven and ten breeding cycles.

The second null hypothesis is tested by simulating a stable population to burn-in, and calculating N_b from the youngest cohort with sampling as in the first hypothesis.

Methods

Details of Simulations.

Using Simupop we simulated structured populations through successive breeding cycles under various scenarios of population decline. Each simulation was set to track 400 single nucleotide polymorphisms (SNPs) for 10 cycles after a burn in period at zero decline. After the burn in period all new members of the population were recorded, and the prescribed decline (as specified later) was imposed by removing a percentage of the individuals from the population.

We determined N_b for each breeding cycle by using the LD method on the young of year (YoY) cohort of each breeding cycle. We then performed linear regression analysis on the N_b values over a period of five, seven and ten years, allowing us to look for significant trends. By replicating this stochastic process, a thousand times for each permutation of the test variables with a known decline in N_c, we experimentally determined how accurately changes in N_c are reflected by changes in N_b.

The LD method was implemented using the program LDNE v.1 (Waples And Do 2008). Using the YoY cohort, the one produced that cycle, maximizes the number of individuals that can be sampled. After calculating N_b we applied a bias adjustment based on the N_b/N_e ratio to increase the accuracy of N_b calculations (Waples 2014).

Both of these programs were integrated into the program AgeStructNe which adds a graphical interface designed by Ted Cosart and other quality of life changes to make these programs easier to use (Hand Unpublished).

Variables Changed During Population Simulation

To determine how changes in N_c are reflected in N_b, simulated a number of different decline situations. To test the specificity of this technique we used a 0% decline (steady population) as a control. Sensitivity to decline was tested by simulating declines at the 5%, 7%, 10%, and 15% level every breeding cycle for 10 cycles. This simulates a range of continuous declines that covers the transition to effectiveness of N_b as a useful measure of decline, as per our investigative tests.

We tested populations starting at 25, 50 and 100 N_b to determine how the different initial N_b levels affected the ability to detect a steady decline. The choice to explore low N_bs was based on 2 factors: that population decline is of much greater interest when N_b is already low, that is when the threat of extinction is high, and that genetic drift has the greatest effect on a population when N_b is low. In bull trout at 25 N_b had insufficient members to sample in the majority of cases

Two species were chosen from the tables verified in Waples et al. (2014). The species were chosen to represent different levels of deviation from the Wright Fisher assumptions, wood frogs (*Lithobates sylvaticus*) best fit the Wright-Fisher assumptions; short lifespan, most of the active breeding population is about one breeding cycle old, and there are equal survival rates for both genders. The other species, bull trout, have a much longer lifespan, and reach sexual maturity at 3 years old. This lets us compare how N_b responds to decline in longer lived species.

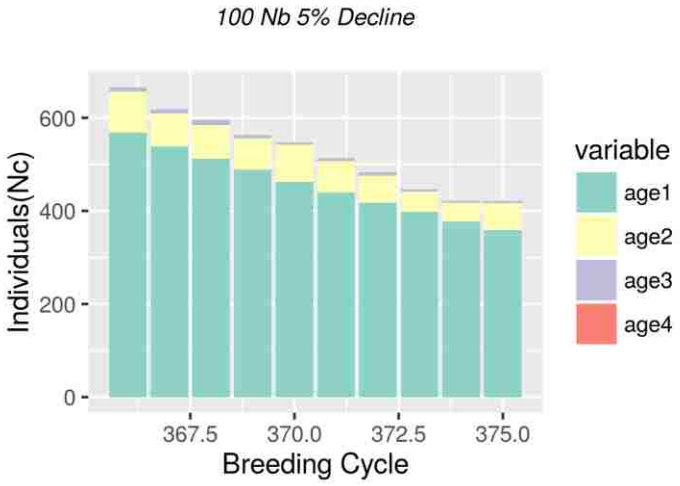


Figure 1:Breakdown of population by age for wood frog during sampling of one simulation under a 5% decline, the first age group (age 1) is the cohort being used to calculate N_b for the population, Age of maturity is 1 breeding cycle.

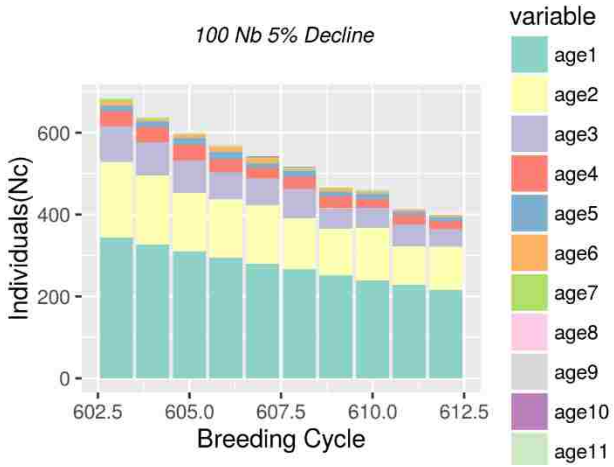


Figure 2:Breakdown of population by age for bull trout during sampling during one simulation under a 5% decline, the first age group (age 1) is the cohort being used to calculate N_b for the population, Age of maturity is 3 breeding cycles.

Of the evolutionary forces that affect allele frequencies, genetic drift is the major contributor and the only one we will consider in our simulations, due to the short timespan monitored. Selection has a limited effect at low effective population size outside of extreme instances of high selective pressure (Frankham 2005).

Because of the lack of mutation, we cannot rely on reaching equilibrium over a significant number of breeding cycles before we start collecting data. To determine an appropriate burn-in

we approximated when heterozygosity of the system reached a level near what is seen in nature. Starting heterozygosity between 0.4 and 0.25 was determined to have little effect on N_b variance (Luikart, unpublished), hence we decided on a heterozygosity of 0.3 for SNPs. How many breeding cycles of burn-in to use was determined by averaging the breeding cycle when the heterozygosity fell below 0.3 for SNPs in sample simulations. Each sample simulation was performed 10 times for each N_b and species combination using 400 SNPS. The critical breeding cycles were recorded, averaged, and rounded down to the nearest integer (Table 1).

Table 1:Table of the number of breeding cycles used for the burn-in period for each species, N_b combination.

Species	25 N_b	50 N_b	100 N_b
Wood Frog	88	180	367
Bull Trout	—	295	604

Calculation of Starting Population Size

.

$$N_0 = \frac{Nb}{Nb/Nc * Nc/N_0} \quad (1)$$

$$Nc/N_0 = \sum_{i=0}^{n}\left(0.5 * \prod_{j=0}^{i} S_m\right) + \sum_{i=0}^{n}\left(0.5 * \prod_{j=0}^{i} S_f\right) \quad (2)$$

To get the desired starting N_b in the simulation we had to calculate values for the starting population size and recruitment class size (N_0). The initial recruitment class size to get a desired N_b, is sum of the cumulative products of the survival tables for male and female by age group adjusted by the N_b/N_c ratio where S_m and S_f are the survival rates for males and females for each age class, respectively (Equation 1 and 2). The Nb/Nc ratios we used were determined by Waples et al.2014 for both species. This gives us a good estimator of the number of new individuals, recruitment class, needed to maintain a specific N_b under ideal conditions. The starting population will be set to twice N_0 rounded up to the nearest 100 to allow for the population to narrow down to the desired population (Table 2).

Table 2: Starting N0 and Starting Individuals for each population.

Species	25 N_b	50 N_b	100 N_b
	(N0, Starting Size)	(N0, Starting Size)	(N0, Starting Size)
Wood Frog	(150, 300)	(299, 600)	(598, 1200)
Bull Trout	—	(181, 300)	(362, 800)

We wanted to consider events whose effects were non age class specific (e.g. habitat loss). Simupop initially makes all the reduction in the recruitment class, leaving the juvenile and adult cohorts intact. We replaced this with a reduction method that performs the percentage reduction equally across all age cohorts including the recruitment class. To do this we separated the population into age and gender classes, then randomly select individuals from each class until that class has been reduced by the desired percentage, stochastically rounded[1].

Sampling Data

To look at how variation in number of samples affected the sensitivity and specificity of detection we looked at sampling all permutations of three different values of SNP loci per individual and three different levels of individual sampling. We calculated N_b for these combinations by subsampling loci and individuals from our simulated data by randomly selecting either 100, 200, or 400 loci and 25, 50, or 100 individuals to use to calculate N_b. For some combinations, there were insufficient members of the recruitment class to sample at the desired level (Table 3). In those cases, the desired value was used until no longer possible then 100% of the cohort was sampled for the rest of the time (Figure 3).

[1] Rounding was performed stochastically by drawing a number from a uniform distribution, then rounding up if that number is greater than the decimal portion, and down if less.

Table 3: Table of Populations that had insufficient members of the YOY cohort to complete sampling.

Species	N_b	Decline	Cycle after burn-in of Insufficient Individuals	Average Individuals after 10 Cycles
	100 Individuals Sampled			
Wood Frog	50	15%	7	58.8
Wood Frog	25	5%	8	89.8
Wood Frog	25	7%	6	72.6
Wood Frog	25	10%	4	52.3
Wood Frog	25	15%	3	29.5
Bull Trout	50	7%	9	87.6
Bull Trout	50	10%	6	63.1
Bull Trout	50	15%	4	35.6
Bull Trout	100	15%	8	71.3
	50 Individuals Sampled			
Wood Frog	25	15%	7	29.5
Bull Trout	50	15%	8	35.6

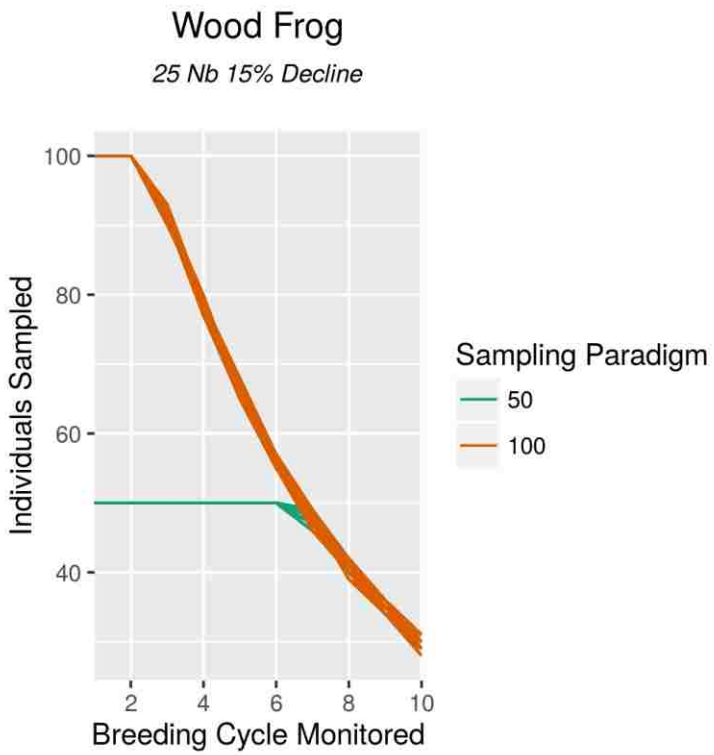


Figure 3:Number of Individuals sampled for each breeding cycle in Wood Frog at 25 starting N_b under a 15% Decline

Performing linear regression on N_b

To determine how accurately changes in population size are reflected by changes in N_b, we must detect changes in N_b, then determine how these changes relate to changes in N_c. Changes in N_b for a set of simulations are determined by performing a least squared linear regression on N_b over the breeding cycles of the sampling time frame. Because the slope of the linear regression is the rate of change of N_b over time, it can be used to evaluate whether to reject the null hypothesis for the slope of the regression at a given significance level.

Breeding Cycles Sampled

Time scale sampling was done by limiting the number of cycles to calculate the linear regression. Starting at the first cycle recorded, the first cycle after our burn-in, we performed a least squares linear regression over the succeeding five, seven, or ten cycles.

Evaluating the Hypothesis for the Slope b_1

To test the significance of the slope of a regression (b_1), we need to determine the degrees of freedom of the linear regression, a test statistic, and the critical p-level. Because the regression is based on N_b not N_c, and we also need to define a hypothesis to test in terms of N_b.

H_o': $b1 = 0$ (neutral slope)

H_a': $b1 \neq 0$ (non- neutral slope)

We tested these hypotheses, with the assumption of a neutral slope. This means we will only determine a population is in decline if we reject this hypothesis. We then can use the results of

this hypothesis to calculate the sensitivity and specificity for our original hypotheses which are in terms of N_c.

Degrees of Freedom

The degrees of freedom for a linear regression is $n - 2$ where n is the number of points used to calculate the regression. Two parameters are estimated in the linear regression, the slope coefficient (b_1) and the intercept (b_0) so the degrees of freedom are reduced by two. Because only the N_b values are used in the regression, and there is one N_b value per cycle, the degrees of freedom for this test is based on the number of cycles monitored. This imposes a minimum limit on the number of sampling events, at least three breeding cycles are needed to perform analysis resulting in one degree of freedom (Neter 1985).

Test Statistic

$$t^* = {b_1}/{s(b_1)} \qquad (3)$$

The test statistic (t*) for the slope of a linear regression can be calculated using the formula above for a normally distributed regression with a null hypothesis equal to zero, where b_1 is the slope of the regression and $s(b_1)$ is an estimate of the variance of the slope (Neter 1985). In most regression models b_1 is assumed to be distributed normally. This assumption is based on the assumptions that the data being regressed over is roughly linear, distributed approximately normal, contains no major outliers, and that the variable being regressed over is independent of the other variables. Least-squared regression is somewhat robust in terms of equal variance and violations of the normality assumption (see figure 4 and 5 to) (Earnst et al. 2017). To calculate $s(b_1)$ requires calculating the mean squared error of the line (Equation 4 and 5). Using t* and the degrees of freedom of the regression we can then calculate the p-value for that line using a Cumulative Density Function (CDF) on the T distribution. (Neter 1985).

$$s(b_1) = \frac{MSE}{\sum(x_i - \bar{x})} \qquad (4)$$

$$\mathrm{MSE} = \frac{\sum(\mathrm{y_i} - \widehat{\mathrm{y_i}})^2}{\mathrm{n} - 2} \qquad (5)$$

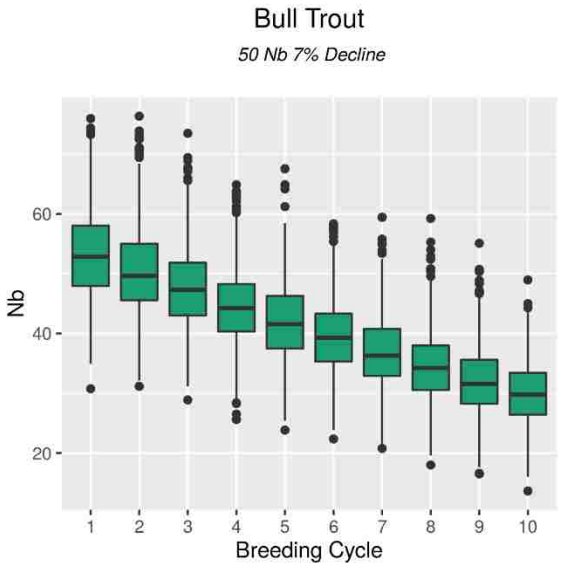


Figure 4: N_b estimates from LDN_e for bull trout starting at 50N_b under 7% decline, sampling 100 individuals and 400 SNPs.

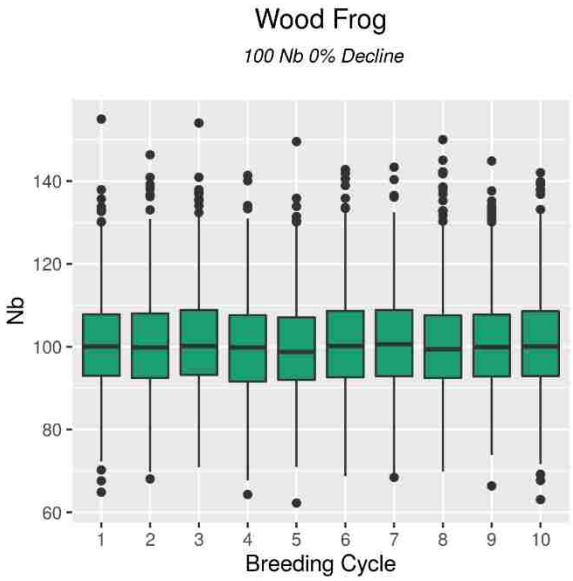


Figure 5: N_b estimates from LDN_e for bull trout starting at 100 N_b under no decline, sampling 100 individuals and 400 SNPs.

P value

The p-value of a hypothesis describes the probability of obtaining the observed result assuming the null hypothesis is true. Because we are looking at the p-value of the linear regression on N_b, we are considering the hypothesis Ho' for the purposes of determining p-value. This lets us set a p-value at which we can reject the assumption of the null hypothesis, essentially how willing we are to say that our assumption is wrong even if it is not. One advantage of using a CDF function is it maintains sidedness allowing differentiation between significant positive and negative values. To determine how this decision affects the tradeoff between sensitivity and specificity of the system, we decided on 3 different p-value cutoffs (85%, 90%, 95%) to analyze how p value corresponds to specificity. In our classification system this p-value is the

discriminating characteristic that determines whether a population size change has been detected (Equation 6).

$$reject\ H_o'\ if: p-value < {}^{p-value\ cutoff}/_{2} \vee 1 - p-value < {}^{p-value\ cutoff}/_{2} \quad (6)$$

ROC Curves

To create ROC curves for a specific permutation of our test factors requires multiple replicates. For each replicate we determined the p-value of that replicate using a CDF function on the t score. Using p-value as a determining variable requires that p-value correlates linearly with specificity. To test for this linearity we will count, for a range from 0 to 1 incrementing by 0.01, the number of simulations that reject H_0' at that p-value cutoff in our control simulations (0% decline) divided by the total number of simulations (Equation 6). This results in a range of 1-specificity (false positives) that will traverse from 0 to 1, keyed by p-value cutoff. See Figure 9 for a plot produced using this methodology.

The steps above are applied to a set of simulation replicates that experienced a decline to get a range of sensitivities. By matching p-value cutoff between the control and the decline sensitivities, we can create a graph of sensitivity over false positives, a ROC curve for each permutation of the test variables. See Figures 10 to 13 for examples of plots produced by this methodology.

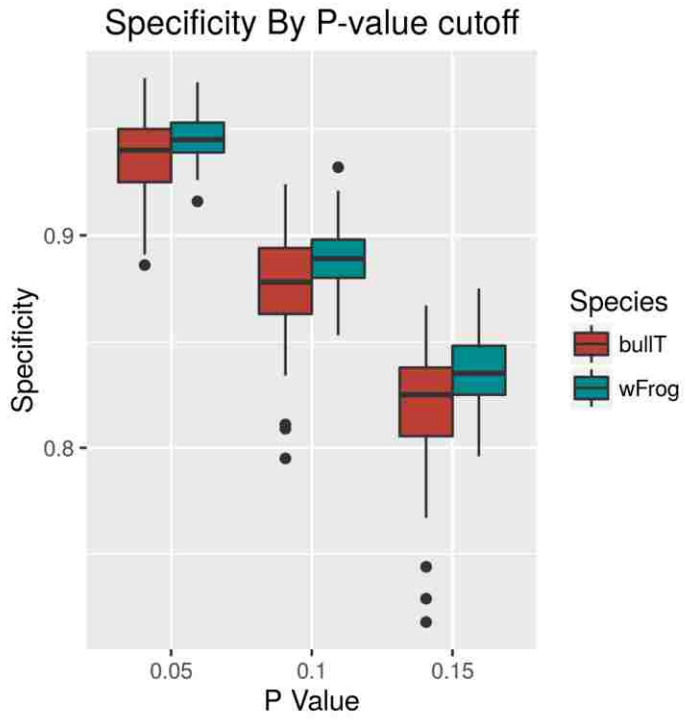
Sensitivity and Specificity

Sensitivity and specificity are a way to measure the accuracy of a system when statistical assumptions do not hold. We will develop a confusion matrix for each system of control and

decline simulations, using 1000 control (no decline) and 1000 decline simulations, these allow us to calculate the sensitivity and specificity of each system of simulations. This gives us a metric to determine how effective using N_b as an indicator of changes in N_c under various conditions. For the purposes of our tests we decided that a sensitivity of 75% and a specificity of 90% would be an acceptable accuracy

Results

Specificity



In both species, we observed, a high negative correlation between specificity and p-value (Figure 6). We also observed a slight increase in variance as we relaxed the p-value. In bull trout the variance in specificity increased as we permutated the factors we tested (Box 1).

Figure 6: Distribution of Specificities (true negative results) compared by species, wood frog (wFrog) and bull trout (bullT), over p-value, combining all other variables tested.

Population Sampling

The combination of both number of individuals and loci have the largest impact on sensitivity (Figure 7and 8), with the number of individuals having a larger effect than the number of loci. Under sampling individuals seems to have the biggest effect on depressing sensitivity, sampling many individuals less than the expected N_b severely diminishes the ability to detect population changes. This is shown by having our lowest sensitivity for both species (2.8% and 2.6%) and the highest (100%) present at the same N_b (Figure 7 and 8, 100 N_b). There is also an interesting result, 400 SNPs sampled has approximately equivalent sensitivity to doubling the individuals sampled at 100 SNPs in nearly all our test cases (Figure 7 and 8).

Breeding Cycles

The number of cycles monitored effected the ability to detect decline. After 5 breeding cycles only, the highest levels of decline tested had an acceptable chance of being detected but only with higher sampling levels (100 Individuals). At 7 breeding cycles we can reliably detect a 10% or greater decline when population sampling is high (Individuals >= expected N_b) except in the case of 25 N_b. After 10 breeding cycles, the ability to detect a decline was significant at all except the smallest (5%) of decline, dependent on the number of SNPs and individuals sampled (Figure 7 and 8).

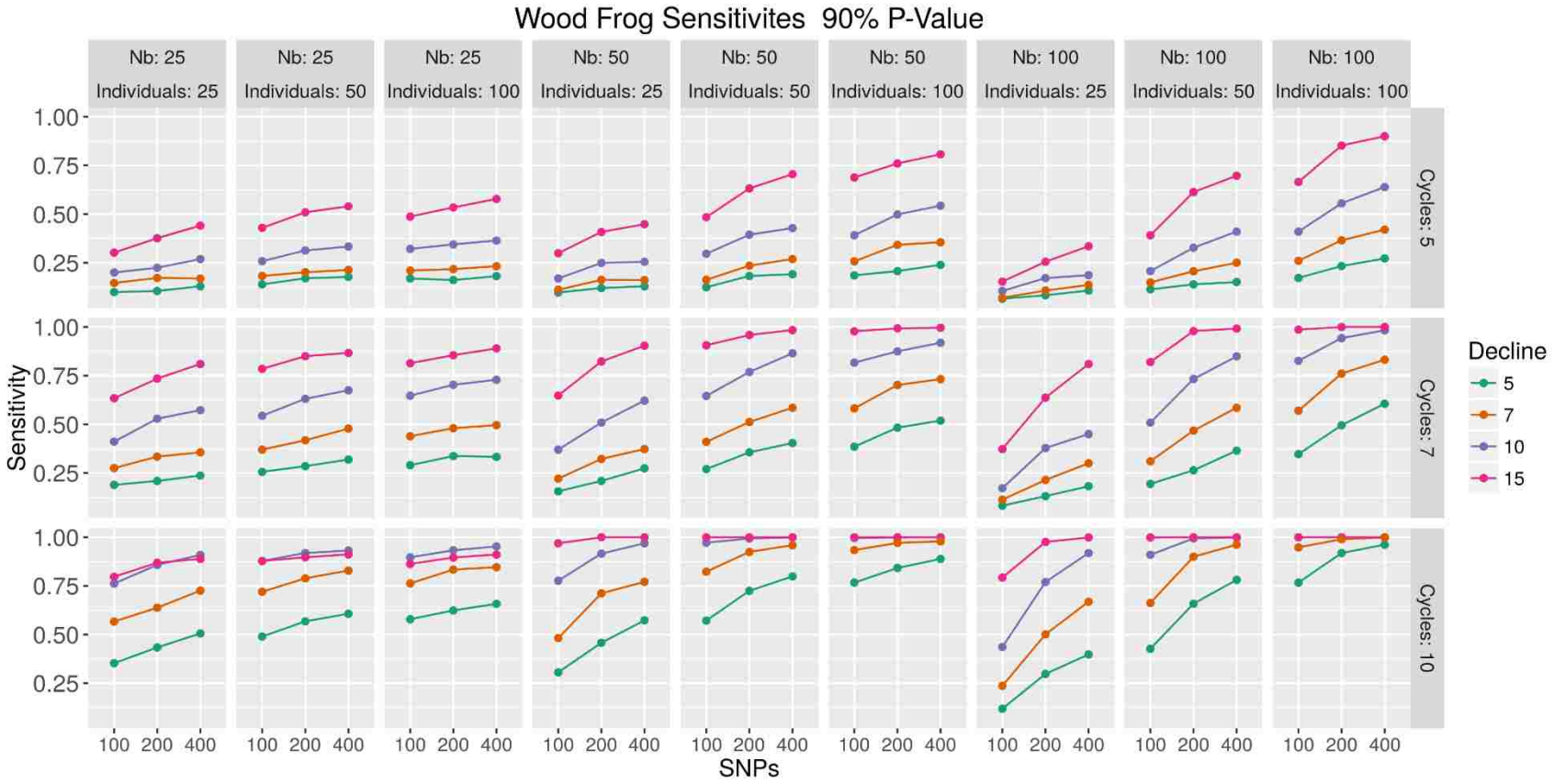


Figure 7: Sensitivities by decline for wood frog at 90% p-value

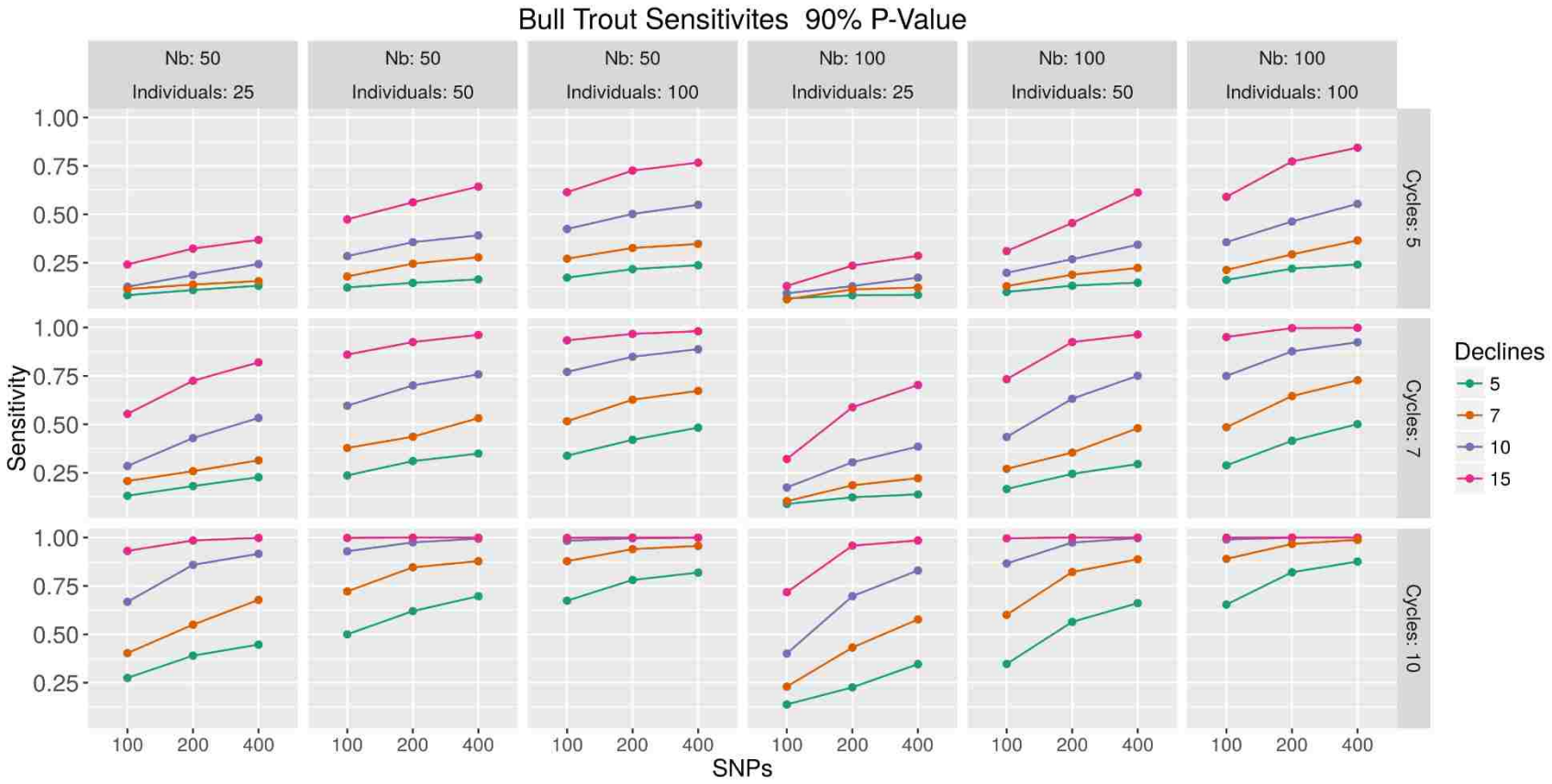


Figure 8: Sensitivities by decline for bull trout at 90% p-value.

P-Value

Looking at the ROC curves allows us to determine how p-value affects sensitivity and specificity. A ROC curve is generated for each permutation of our test variables, these can be used to evaluate the tradeoffs between sensitivity and specificity as we vary our determining characteristic, p-value for slope of the N_b regression. The first observance is the linearity of the graph, indicating that specificity is highly correlated to p-value (Figure 9). We can use these false positive values to create ROC curves for sensitivity. Because of the linearity of false positives to p-value these ROC curves accurately reflect how monitoring at different p-values will affect both sensitivity and specificity at the various decline levels. By plotting the ROC curves for each parameterization of our test factors we can compare how sensitivity reacts to the different p values at different levels of decline Figures (10-13). The closer the curve approaches the upper left corner (0,1) of the graph the more accurate that combination of variables is at detecting both stability and changes in the population (i.e. high sensitivity and specificity) (Figures 12 and 13), while a curve that approaches the line $y = x$ means there is little difference between that detection method and deciding randomly (i.e. low sensitivity and specificity) (Figure 10).

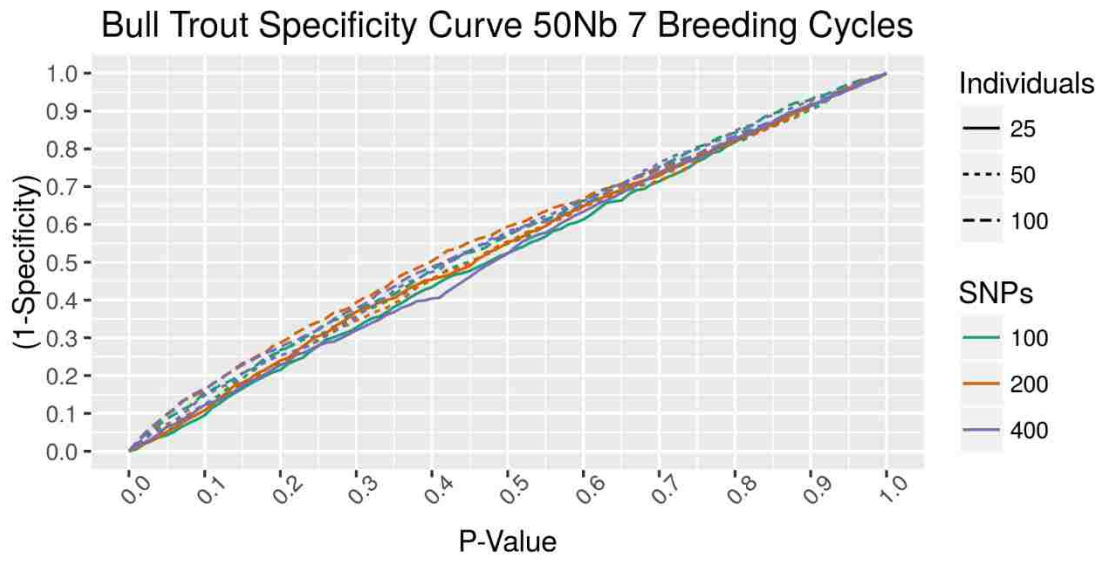


Figure 9:Specificity Curve for bull trout at 50N_b, when sampling 100, 50 or 25 individuals from the youngest cohort and sampling 100, 200 0r 400 SNPs per individual.

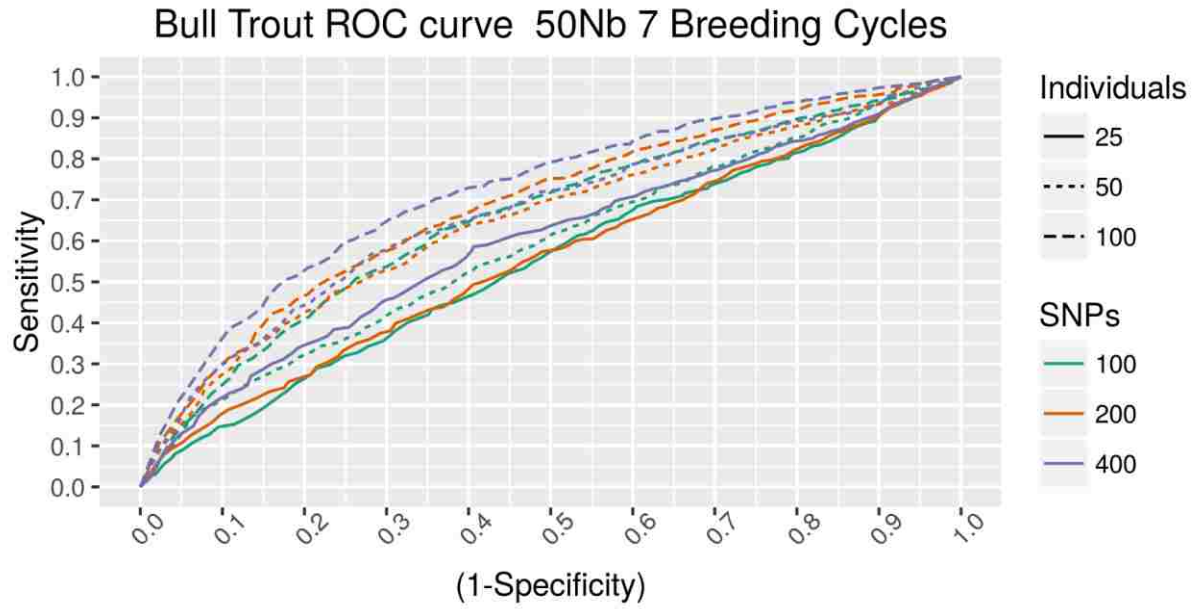


Figure 10: ROC Curve for Individuals and Loci for bull trout at 50N_b and 7 breeding cycles under a 5% decline per year, when sampling 100, 50 or 25 individuals from the youngest cohort and sampling 100, 200 0r 400 SNPs per individual.

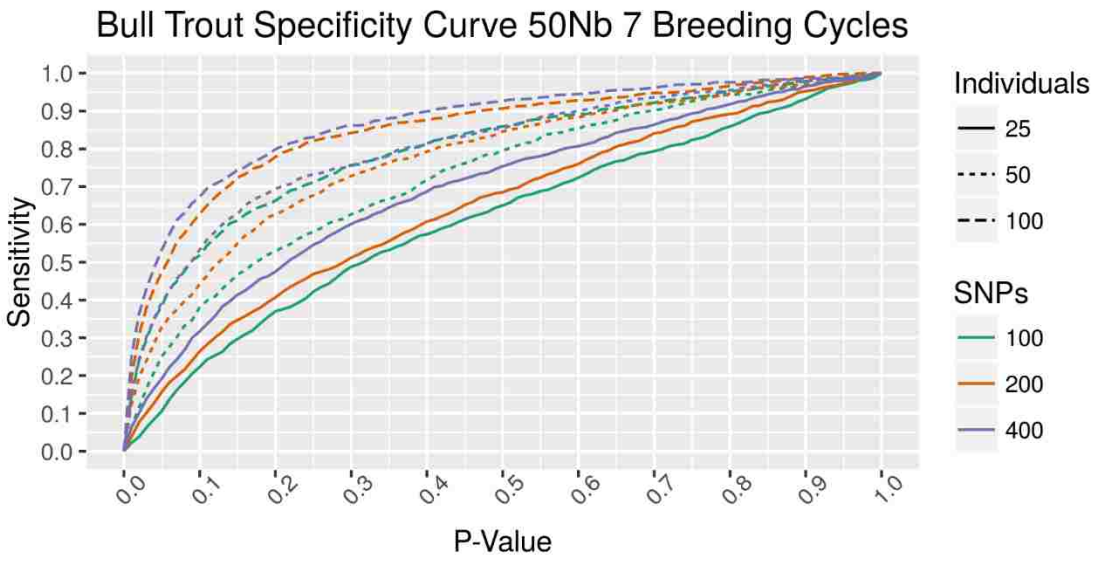


Figure 11: ROC Curve for Individuals and Loci for bull trout at 50N_b and 7 breeding cycles under a 7% decline per year, when sampling 100, 50 or 25 individuals from the youngest cohort and sampling 100, 200 0r 400 SNPs per individual.

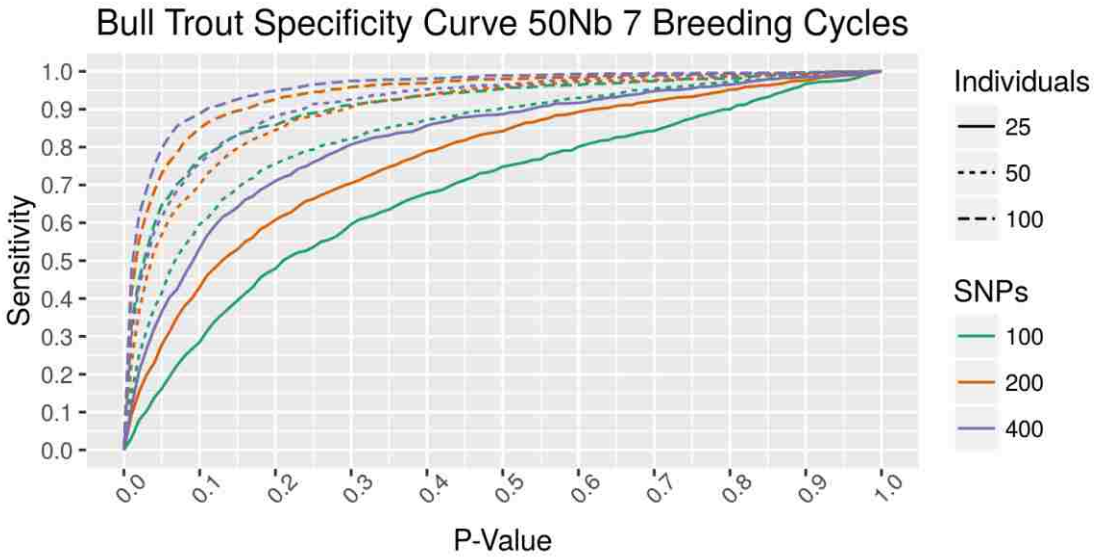


Figure 12:ROC Curve for Individuals and Loci for bull trout at 50N_b and 7 breeding cycles under a 10% decline per year, when sampling 100, 50 or 25 individuals from the youngest cohort and sampling 100, 200 0r 400 SNPs per individual.

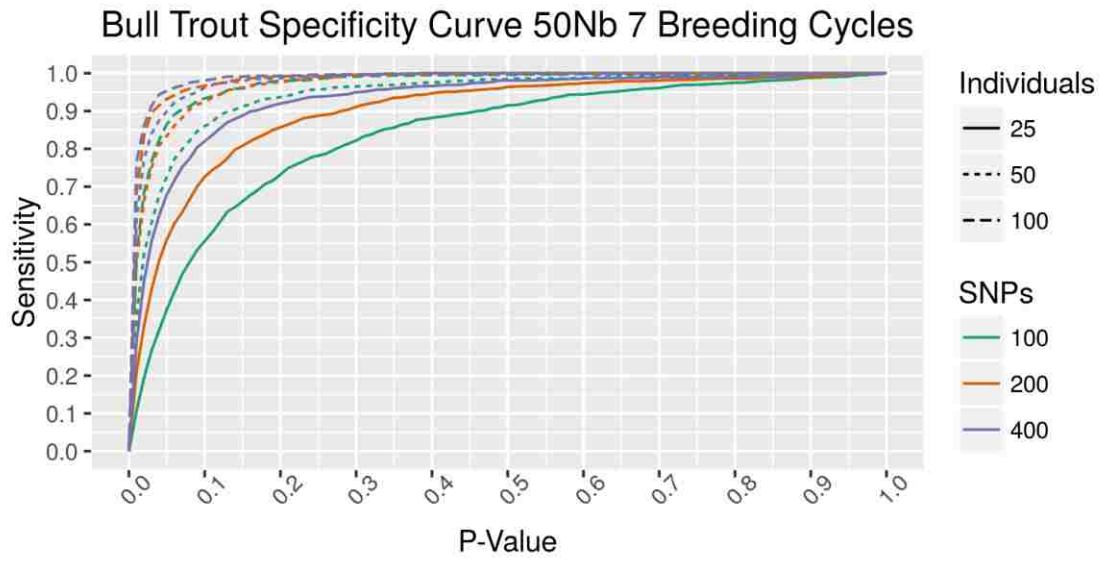


Figure 13:ROC Curve for Individuals and Loci for bull trout at 50N_b and 7 breeding cycles under a 15% decline per year, when sampling 100, 50 or 25 individuals from the youngest cohort and sampling 100, 200 0r 400 SNPs per individual.

Discussion

Using N_b as an indicator of population declines was very successful under certain conditions, including after long periods of decline (10 Breeding Cycles), and when many individuals and loci were sampled (100 individuals and 400 SNPs). We also identified three cases where N_b failed as an indicator (had a sensitivity less than 75%) for anything but the largest declines (declines > 15%). When N_b was low (N_b = 25), when monitoring time was short (5 breeding cycles), and when insufficient individuals were sampled in comparison to N_b (e.g. N_b > number of individuals sampled).

The major variables of the study can be separated into 3 categories: those that are observed during testing and may affect sensitivity and specificity (N_b and decline), those that can be controlled (Individuals sampled, SNPs sampled, and number of cycles monitored), and those that are determined during the study design period (target specificity and sensitivity).

The first set of variables, those that are unknown during study design, must be estimated by some measure to determine acceptable ranges for the other variables. An estimate of N_b helps set a minimum appropriate number of individuals to sample, and an estimate is needed for the minimum decline to identify to determine the correct p-value to use as a determining factor.

In general sensitivity increases with an increased number of individuals and SNPs sampled, with increasing the number of individuals having a larger effect on sensitivity. One noteworthy result was that by sampling at least as many individuals as your estimated N_b prevents under sampling, which causes higher variation in N_b estimations using LDN_e and leads to lowered sensitivity. This observation has been corroborated by Ackerman in steelhead (*Oncorhynchus mykiss*) hatchery fish (Ackerman et al. 2016). Increasing individuals sampled also had a larger effect on sensitivity than increasing SNPs.

Sensitivity also increases with the number of breeding cycles monitored. After 7 cycles a 7% or greater decline was detectable in some cases, this increased to being able to detect a 5% or greater decline after 10 breeding cycles. This can make monitoring challenging in populations with especially low N_b, because in the event of a slow steady decline, it can be difficult to detect in time to implement population saving management strategies.

The last set of variables are specific to the study. The acceptable rates for sensitivity and specificity are defined in a large part by an analysis based on the situation and the consequences of acting on incorrect information. For example, when a recovery program to stabilize the species of interest is expensive or requires significant regulation to protect habitat, it would be safer to err on the side of concluding the population is stable even if it not. This would require a higher specificity to limit the chances of a false positive result and therefore a

higher p-value. On the other hand, if losing the population in question would be significant, such as in the case of near extinction, preservation would be of greater importance and therefore a higher sensitivity would be desired, necessitating a lower p-value. These types of concerns should be determined before the study begins. These target sensitivities and specificities can be used to determine the p-value through use of ROC curves. Contingent on specificities strong relationship to p-value (Figure 4) and the strong linearity seen between false positive rate ($1 - specificty$) and p-value (Figure 7), this indicates that desired false positive rate ($1 - target\ specificity$) is a good estimator of which p-value to use. To determine if that specificity is appropriate for a set of parameters a ROC curve should be created to see if the target sensitivity is possible for that combination of variables.

While using ROC Curves to decide on a specificity and sensitivity it should be noted that, a high specificity with a low sensitivity appears suitable for detecting stability in population size however, there could, in fact, be a high rate of false negative results despite the population being in decline.

Some thought should be given to the life history and breeding patterns of the species being studied. The two species tested showed minimal difference in sensitivity and specificity, but this is no indication how other violations of the Wright-Fisher model, such as unequal gender ratios or nonrandom mating will affect sensitivity and specificity.

Additional Studies

This study is just a preliminary exploration of how N_b can be used to monitor populations for decline. Further exploration into how life characteristics affect this methodology can be tested by expanding the number of species tested, especially if the additional species further stretch

assumptions of the Wright-Fisher model, such as species with unequal survival rates between the genders or different mating habits (e.g. monogamous family mating) would greatly increase the viability of this method in real world situations. Another area of expansion would be exploring different decline scenarios, such as how easy it is to detect large one-time decline events or periodic declines with either periods of stability or growth in between. How population size is reduced (e.g. reduction in only the adults or juveniles before the age of maturity) is another area that has much potential for exploration.

References

Ackerman, M. W., Hand, B. K., Waples, R. K., Luikart, G., Waples, R. S., Steele, C. A., ... & Campbell, M. R. (2017). Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. *Evolutionary applications*, *10*(2), 146-160.

Bernos, T. A., & Fraser, D. J. (2016). Spatiotemporal relationship between adult census size and genetic population size across a wide population size gradient. *Molecular ecology*, *25*(18), 4472-4487.

Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature reviews. Genetics*, *10*(3), 195.

Cornuet, J. M., & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, *144*(4), 2001-2014.

Collen, B., McRae, L., Deinet, S., De Palma, A., Carranza, T., Cooper, N., ... & Baillie, J. E. (2011). Predicting how populations decline to extinction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1577), 2577-2586.

Ernst, A. F., & Albers, C. J. (2017). Regression assumptions in clinical psychology research practice—a systematic review of common misconceptions. *PeerJ*, *5*, e3323.

Frankham, R. (2005). Genetics and extinction. *Biological conservation*, *126*(2), 131-140.

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., ... & Palstra, F. (2011). Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology*, *25*(3), 438-449.

Hartl, D. L., Clark, A. G., & Clark, A. G. (2006). *Principles of population genetics* Fourth Edition Sinauer associates.

Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *TAG Theoretical and Applied Genetics*, *38*(6), 226-231.

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics*, *11*(2), 355-373.

Mace, G. M., Collar, N. J., Gaston, K. J., Hilton- Taylor, C. R. A. I. G., Akçakaya, H. R., Leader- Williams, N. I. G. E. L., ... & Stuart, S. N. (2008). Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, *22*(6), 1424-1

Neter, J., Kutner, M. H., & Wasserman, W. (1985). *Applied linear statistical models: regression, analysis of variance, and experimental designs* (2nd ed.). Homewood, IL: Irwin.

Peng, B., & Amos, C. I. (2008). Forward-time simulations of non-random mating populations using simuPOP. *Bioinformatics*, *24*(11), 1408-1409.

Peng, B., & Kimmel, M. (2005). simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, *21*(18), 3686-3687.

Perrier, C., April, J., Cote, G., Bernatchez, L., & Dionne, M. (2016). Effective number of breeders in relation to census size as management tools for Atlantic salmon conservation in a context of stocked populations. *Conservation genetics*, *17*(1), 31-44.

Tallmon, D. A., Gregovich, D., Waples, R. S., Scott Baker, C., Jackson, J., Taylor, B. L., ... & Schwartz, M. K. (2010). When are genetic methods useful for estimating contemporary abundance and detecting population trends?. *Molecular Ecology Resources*, *10*(4), 684-692.

Tallmon, D. A., Waples, R. S., Gregovich, D., & Schwartz, M. K. (2012). Detecting population recovery using gametic disequilibrium-based effective population size estimates. *Conservation Genetics Resources*, *4*(4), 987-989.

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, *7*(2), 167-184.

Waples, R. S., & Do, C. H. I. (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular ecology resources*, *8*(4), 753-756.

Waples, R. S., Do, C., & Chopelet, J. (2011). Calculating N_e and N_e/N in age- structured populations: a hybrid Felsenstein- Hill approach. *Ecology*, *92*(7), 1513-1522.

Waples, R. S., Luikart, G., Faulkner, J. R., & Tallmon, D. A. (2013). Simple life-history traits explain key effective population size ratios across diverse taxa. *Proceedings of the Royal Society of London B: Biological Sciences*, *280*(1768), 20131339.

Waples, R. S., Antao, T., & Luikart, G. (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics*, *197*(2), 769-780.

Whiteley, A. R., Coombs, J. A., Cembrola, M., O'Donnell, M. J., Hudy, M., Nislow, K. H., & Letcher, B. H. (2015). Effective number of breeders provides a link between interannual variation in stream flow and individual reproductive contribution in a stream salmonid. *Molecular ecology*, *24*(14), 3585-3602.

Whiteley, A. R., Coombs, J. A., O'Donnell, M. J., Nislow, K. H., & Letcher, B. H. (2017). Keeping things local: Subpopulation N_b and N_e in a stream network with partial barriers to fish migration. Evolutionary Applications, 10(4), 348-365.

Yuan, X., Miller, D. J., Zhang, J., Herrington, D., & Wang, Y. (2012). An overview of population genetic data simulation. *Journal of Computational Biology*, *19*(1), 42-54.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, *19*.