

RESOURCE ARTICLE OPEN ACCESS

MAXTEMP: A Method to Maximise Precision of the Temporal Method for Estimating N_e in Genetic Monitoring Programs

Robin S. Waples¹  | Michele M. Masuda² | Melanie E. F. LaCava³  | Amanda J. Finger³

¹School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington, USA | ²Auke Bay Laboratories, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Juneau, Alaska, USA | ³Department of Animal Science, University of California Davis, Davis, California, USA

Correspondence: Robin S. Waples (robinw3@uw.edu)

Received: 3 June 2024 | **Revised:** 2 December 2024 | **Accepted:** 17 December 2024

Handling Editor: Frederic Austerlitz

Funding: Genetic work for delta smelt was supported through Bureau of Reclamation Grants #R10AC20089, #R15AC00030 and #R20AC00027, and State Water Contractors Agreement A19-1844.

Keywords: computer simulations | effective population size | genetic drift | genetic monitoring | linkage disequilibrium | temporal method

ABSTRACT

We introduce a new software program, *MAXTEMP*, that increases precision of the temporal method for estimating effective population size (N_e) in genetic monitoring programs, which are increasingly used to systematically track changes in global biodiversity. Scientists and managers are typically most interested in N_e for individual generations, either to match with single-generation estimates of census size (N) or to evaluate consequences of specific management actions or environmental events. Systematically sampling every generation produces a time series of single-generation estimates of temporal F (\hat{F}), which can then be used to estimate N_e ; however, these estimates have relatively low precision because each reflects just a single episode of genetic drift. Systematic sampling also produces an array of multigenerational temporal estimates that collectively contain a great deal of information about genetic drift that, however, can be difficult to interpret. Here, we show how additional information contained in multigenerational temporal estimates can be leveraged to increase precision of \hat{F} for individual generations. Using information from one additional generation before and after a target generation can reduce the standard deviation of \hat{F} ($\sigma_{\hat{F}}$) by up to 50%, which not only tightens confidence intervals around \hat{N}_e but also reduces the incidence of extreme estimates, including infinite estimates of N_e . Practical application of *MAXTEMP* is illustrated with data for a long-term genetic monitoring program for California delta smelt. A second feature of *MAXTEMP*, which allows one to estimate N_e in an unsampled generation using a combination of temporal and single-sample estimates of N_e from sampled generations, is also described and evaluated.

1 | Introduction

Effective population size (N_e) is one of the most important concepts in evolutionary biology but also one of the most enigmatic (Charlesworth 2009; Hare et al. 2011; Waples 2022). Because it is challenging to collect the demographic data from natural populations that is required to estimate N_e directly, for over half a century researchers have used genetic methods to indirectly

estimate N_e . Initially most of these estimates used the temporal method, which compares allele frequencies in samples from the same population taken at different points in time (Krimbas and Tsakas 1971; Nei and Tajima 1981; Pollak 1983; Waples 1989; Wang 2001). That changed abruptly in the late 2000s with development of two new estimators that require only a single sample: a method based on linkage disequilibrium, LD (Waples 2006; modified from Hill 1981) and a method

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Published 2025. This article is a U.S. Government work and is in the public domain in the USA. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

based on the incidence of siblings (Wang 2009). Within a few years, use of single-sample estimators had far outstripped the vintage temporal method (Palstra and Fraser 2012). A recent review (Clarke et al. 2024) conducted a meta-analysis of over 4600 estimates of N_e from the LD method alone. This trend toward increasing interest in estimating N_e is likely to become even stronger in the future. Implementation of regular genetic monitoring has increased in recent decades along with the realisation that global ecosystems are undergoing rapid changes (Schwartz, Luikart, and Waples 2007; De Barba et al. 2010; Foote et al. 2012; Jackson et al. 2012; Fussi et al. 2016; Van Rossum and Hardy 2022). Furthermore, N_e has been identified as a key metric to monitor with respect to the Convention on Biological Diversity's post-2020 global biodiversity framework (Hoban et al. 2022; Thurfjell et al. 2022). Given the realisation that global ecosystems are undergoing rapid changes, the trend toward increasing interest in estimating N_e is expected to continue.

These are encouraging developments, but it is unfortunate that the full potential of the temporal method is not being realised. A major feature of the temporal method is that it integrates information about effective size across all the generations spanned by the samples, such that the resulting estimate applies to the harmonic mean N_e over the intervening generations. Each generation of genetic drift strengthens the overall signal of effective size, so (up to a point) precision increases with elapsed time between samples. In applied conservation biology and management, however, one often wants to be able to estimate N_e for specific individual generations, either to match with single-generation estimates of census size (N) or to help evaluate consequences of specific management actions or environmental

events (Kamath et al. 2015; Whiteley et al. 2015; Ruzzante et al. 2016). Single-sample estimates apply to individual generations and hence are well suited to meet these needs. If samples are taken in consecutive generations, the temporal method can also provide estimates of N_e that apply to specific generations. In that case, however, a single generation of genetic drift provides a relatively weak signal for estimating N_e , so power is generally less than for single-sample estimators.

Here, we show that in a genetic monitoring program that involves systematic sampling, precision of the standard temporal method to estimate N_e in specific generations can be increased by mobilising information contained within the many multigeneration temporal estimates that also can be made. To illustrate, consider a genetic monitoring program that takes samples of progeny from 4 consecutive generations (Figure 1). Comparison of allele frequencies in samples of progeny from generations 1 and 2 provides an estimate of variance N_e in generation 2 (\hat{N}_{e2}), comparison of samples from generations 2 and 3 leads to \hat{N}_{e3} , and so on, where the '^' indicates an estimate. In addition, comparison of samples from generations 1 and 3 provides an estimate of the harmonic mean N_e in generations 2 and 3:

$$\tilde{N}_{e2-3} = \frac{2}{1/\hat{N}_{e2} + 1/\hat{N}_{e3}}, \quad (1)$$

where the tilde (~) indicates a harmonic mean and the asterisk (*) indicates an estimate that is part of a combined (multigeneration) estimate. Consider now the true effective size in generation 3, which we can estimate directly using samples 2 and 3. We have a second estimate of N_e in generation 3, embedded in the

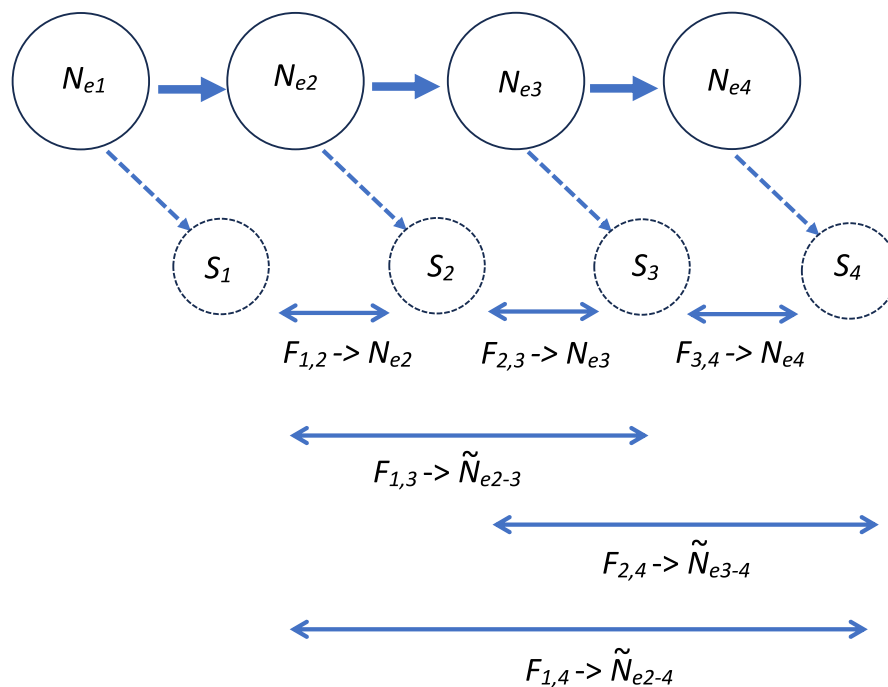


FIGURE 1 | Schematic of 4 consecutive generations of samples used to estimate effective size using the standard temporal method. Plan II samples of S progeny are taken from each generation. Resulting temporal F values estimate single-generation N_e if the samples are separated by one generation. Samples separated by more than a single generation estimate harmonic mean N_e in the generations between samples. The long arrows at the bottom depict all temporal comparisons spanning multiple generations that include a drift signal from N_{e3} . See Figure S1 for an expansion of this model for a scenario involving samples in 5 consecutive generations.

right side of Equation (1). We also have a single-generation temporal estimate of N_e in generation 2, and if we substitute \hat{N}_{e2} for \hat{N}_{e2*} in Equation (1) and rearrange,

$$1/\hat{N}_{e3*} = \frac{2}{\hat{N}_{e2-3}} - 1/\hat{N}_{e2} \quad (2)$$

we can express the second estimate from generation 3 as a function of two temporal estimates. An improved estimate for generation 3 can then be obtained by properly weighting \hat{N}_{e3} and \hat{N}_{e3*} . Furthermore, this process can be extended to extract more information about N_e in generation 3 from the temporal estimate \hat{N}_{e2-4} , which estimates harmonic mean N_e in generations 3 and 4 (Figure 1). Whereas \hat{N}_{e3*} as defined above leverages information about drift in generation 3 contained in the multigenerational estimate \hat{N}_{e2-3} , which is confounded by a drift signal from the generation before generation 3, the temporal estimate \hat{N}_{e2-4} is confounded by a drift signal from the generation after generation 3. Properly weighting all three estimates (the direct estimate \hat{N}_{e3} and the before- and after- versions of \hat{N}_{e3}) then can lead to an improved overall estimate for the focal generation (3 in this case).

In theory, this approach can be extended to arbitrarily long time series. With n consecutive samples spanning $n-1$ generations of genetic drift, there are $\sum_{i=1}^{n-1} i = n(n-1)/2$ different temporal comparisons that provide information about N_e . For example, with $n=10$ consecutive samples, $10 \times 9/2 = 45$ different temporal comparisons contain information relevant to N_e in the 9 generations spanned by the samples. However, we expect the information content to dwindle as the signal from N_e in the focal generation becomes confounded with drift signals from additional generations.

Here we use simulations to explore the feasibility of this idea. We introduce the software *MAXTEMP*, which implements our results to maximise precision for temporal estimates of N_e in individual generations. We show that the percentage reduction in the standard deviation of the single-generation temporal estimator ($\sigma_{\hat{F}}$) is a function of three covariates (true N_e , and sample sizes of individuals (S) and loci (L)) and that reductions of up to about 50% can be achieved with values of S and L that are widely used. We also illustrate a second feature of *MAXTEMP*, which deals with incomplete sampling. In a long-term study there are often data gaps, which (despite the best intentions) can arise for a variety of reasons (e.g., lapse in grant support; departure of key personnel; freezer meltdown; random screw-ups; global pandemic). Consider the above scenario but without the sample from generation 2. The temporal estimate \hat{N}_{e2-3} , based on samples from generations 1 and 3, provides an estimate of harmonic mean N_e in generations 2 and 3. If a single-sample method can be used to obtain \hat{N}_{e3} for generation 3, that single-generation estimate can be used in Equation (2) to obtain an estimate of N_e applicable directly to generation 2.

2 | Methods

Notation used in this paper is defined in Table 1.

TABLE 1 | Notation.

N_e	Effective population size for one or more generations
N_b	Effective number of breeders in 1 year or season
N	Census size
S	Number of individuals sampled
L	Number of diallelic (SNP) loci
L'	Effective number of loci, after accounting for lack of independence due to physical linkage
\wedge	Denotes an estimate
\sim	Denotes a harmonic mean
$E()$	Denotes an expectation
N_{ex}	N_e for generation x
\hat{N}_{ex-y}	An estimate of harmonic mean N_e for generations x through y , inclusive
\hat{N}_{ex*}	An estimate of N_e in generation x that is embedded in a multigenerational temporal estimate that spans generation x (as in Equation 1)
Mystery N_e	N_e in an unsampled generation, which can be inferred by joint use of temporal and single-sample methods as described here
\hat{N}_{exLD}	An estimate of N_e in generation x based on the LD method
t	Number of generations between temporal samples
P	Allele frequency
F	An index of allele frequency change over time
\hat{F}	An estimator of parametric F
F'	Raw F adjusted for random sampling error
$\hat{F}_{x,y}$	An estimate of F based on samples taken in generations x and y
\hat{F}^{Adj}	\hat{F} after adjustment to improve precision using methods described here
\hat{N}_e^{Adj}	An estimate of N_e based on \hat{F}^{Adj}
$\sigma_{\hat{F}}$	The standard deviation of \hat{F}
$\sigma_{\hat{F}^{Adj}}$	The standard deviation of \hat{F}^{Adj}
Ratio	$\sigma_{\hat{F}^{Adj}} / \sigma_{\hat{F}} =$ an index that quantifies improvement to precision
W	A weight used in constructing the optimal \hat{F}^{Adj}
Target N_e	In simulations, a fixed N_e value that was allowed to vary in certain generations

2.1 | Temporal Estimation of N_e

The temporal method for estimating N_e relies on the fact that the standardised variance in allele frequencies comparing two samples separated by t generations of genetic drift (tempF, hereafter just F) is a simple function of N_e and the sample sizes of individuals at times 0 and t (S_0, S_t) (Nei and Tajima 1981; Waples 1989):

$$E(F_{0,t}) = \frac{1}{2S_0} + 1 - \left(1 - \frac{1}{2S_t}\right) \left(1 - \frac{1}{2N_e}\right)^t \quad (3)$$

where $E(F_{0,t})$ refers to the expected value of F calculated for samples of progeny from generations 0 and t . Equation (3) applies to sampling Plan II of Nei and Tajima (1981) and Waples (1989), which requires that the initial sample of individuals that are progeny of reproduction in generation 0 cannot be involved in reproduction in generation 1. If any individuals in the initial sample can contribute genes to subsequent generations, then sampling is according to Plan I and to account for correlations of allele frequencies over time, it is necessary to add a term to Equation (3) for the total number of individuals subject to sampling at the time of the initial sample. We don't attempt that here, but in theory it could be done subsequently.

When elapsed time is small compared to N_e ($t/N_e \ll 1$), F increases approximately linearly by magnitude $1/(2N_e)$ per generation (Nei and Tajima 1981; Mwima et al. 2024), and a good approximation is:

$$E(F_{0,t}) \approx \frac{t}{2N_e} + \frac{1}{S} \quad (4)$$

Rearrangement of Equation (4) produces an estimator of N_e :

$$\hat{N}_e \approx \frac{t}{2\left(F - \frac{1}{S}\right)} = \frac{t}{2F'}, \quad (5)$$

where $F' = F - 1/S$ is the raw temporal F adjusted for sampling a finite number of individuals and $1/S = 1/(2S_0) + 1/(2S_t)$. If S differs between the two samples, the harmonic mean of the two sample sizes should be used for S in the above equations. Luikart, Cornuet, and Allendorf (1999) suggested a modification to Equation (5) if a severe bottleneck affects linearity of the relationship between F and N_e .

The above equations apply to a constant N_e and need to be modified to allow N_e to vary over time. Assume now that the true N_e values for each generation are labelled $N_{e1}, N_{e2},$ etc. Then Equation (3) can be modified as

$$E(F_{0,t}) = \frac{1}{2S} + 1 - \left(1 - \frac{1}{2S}\right) \left(1 - \frac{1}{2N_{e1}}\right) \left(1 - \frac{1}{2N_{e2}}\right) \dots \left(1 - \frac{1}{2N_{et}}\right), \quad (6)$$

which can be approximated by

$$\hat{F}_{2,3}^{\text{Adj}} = \left[W\hat{F}_{2,3} + W_A\hat{F}_{2,3(A)} + W_B\hat{F}_{2,3(B)} + W_C\hat{F}_{2,3(C)} + W_D\hat{F}_{2,3(D)} + W_E\hat{F}_{2,3(E)} \right] / \Sigma W \quad (10)$$

$$E(F_{0,t}) \approx \frac{1}{S} + \frac{1}{2N_{e1}} + \frac{1}{2N_{e2}} \dots + \frac{1}{2N_{et}}. \quad (7)$$

For each of the single-generation N_e terms in Equation (7), we can substitute the respective value for F spanning a single generation:

$$E(F_{0,t}) \approx \frac{1}{S} + F_{0,1} + F_{1,2} \dots + F_{t-1,t} \quad (8)$$

This is very useful, as the distribution of \hat{F} is well-known and well-behaved (Lewontin and Krakauer 1973; Nei and Tajima 1981), whereas the distribution of \hat{N}_e is highly skewed and can take biologically impossible (infinitely large) values. Hereafter we deal with \hat{F} for estimation and only convert to \hat{N}_e after all adjustments to \hat{F} have been made.

Now focus on estimating N_e in a single generation, using the example in Figure S1. Let's say we want to use all available information to estimate N_{e3} . The 5 generations of samples can be used to generate a triangular matrix of $\hat{F}_{x,y}$ values spanning 1–4 generations of genetic drift (Table S1). The most direct way to estimate N_{e3} is to calculate $\hat{F}_{2,3}$ for samples of progeny from generations 2 and 3, as that estimator is not affected by a drift signal from any other generations. But there is also information about N_{e3} in other pairs of samples that span generation 3: samples 1,3; 1,4; 1,5; 2,4; and 2,5. Consider the samples from generations 1 and 3, which produce the estimator $\hat{F}_{1,3}$. From Equation (8),

$$E(F_{1,3}) \approx \frac{1}{S} + F_{1,2} + F_{2,3}$$

Solving for the drift signal specific to generation 3 produces another temporal estimator:

$$\hat{F}_{2,3(A)} = \hat{F}_{1,3} - \hat{F}_{1,2} - \frac{1}{S} \quad (9)$$

This same procedure can be used to generate a full series of estimators applicable to N_{e3} :

$$\hat{F}_{2,3(B)} = \hat{F}_{1,4} - \hat{F}_{1,2} - \hat{F}_{3,4} - \frac{1}{S}$$

$$\hat{F}_{2,3(C)} = \hat{F}_{1,5} - \hat{F}_{1,2} - \hat{F}_{3,4} - \hat{F}_{4,5} - \frac{1}{S}$$

$$\hat{F}_{2,3(D)} = \hat{F}_{2,4} - \hat{F}_{3,4} - \frac{1}{S}$$

$$\hat{F}_{2,3(E)} = \hat{F}_{2,5} - \hat{F}_{3,4} - \hat{F}_{4,5} - \frac{1}{S}$$

To get an overall estimate, we give each separate estimator a weight (W) that is the reciprocal of its variance. Then, the overall temporal F estimator for N_e in generation 3 is

where ‘Adj’ indicates an adjusted estimate that includes information from additional generations. The overall estimator of N_{e3} is then

$$\hat{N}_{e3}^{\text{Adj}} = \frac{1}{2 \hat{F}_{2,3}^{\text{Adj}}}$$

2.2 | Precision

Temporal F has a distribution that is a multiple of a χ^2 distribution (Nei and Tajima 1981; Pollak 1983), but the distribution becomes effectively normal if at least 50 independent alleles are used to compute \hat{F} . Here, we consider diallelic (SNP) loci, which have one independent allele per locus, and we assume that enough loci are used that \hat{F} can be considered to be normally distributed. This means that Wald confidence intervals (CIs) for \hat{F} can be calculated as a simple function of the standard deviation of \hat{F} , $\sigma_{\hat{F}}$. For example, the 95% CI for \hat{F} would be

$$95\% \text{ CI for } \hat{F} = \left[\hat{F} - 1.96\sigma_{\hat{F}}, \hat{F} + 1.96\sigma_{\hat{F}} \right] \quad (11)$$

The CIs of \hat{F} for individual generations are calculated independently, without regard to serial dependence of \hat{F} . CIs for \hat{F} are then used as in Equation (5) to generate CIs for \hat{N}_e .

Under pure drift, the expected variance of \hat{F} for L diallelic loci is $E(\sigma_{\hat{F}}^2) = 2\hat{F}^2 / L$ (Lewontin and Krakauer 1973), so the expected (parametric) standard deviation is $\sigma_{\hat{F}} = \sqrt{2\hat{F}^2 / L}$. This provides a useful reference point for evaluating increases in precision obtained by using more of the information contained in multi-generation temporal estimates. Using computer simulations (described below), we empirically measured $\sigma_{\hat{F}}$ across replicates for single-generation temporal estimates and compared that with $\sigma_{\hat{F}}$ for adjusted estimates that also used information from one or more multigeneration temporal estimates that spanned the focal generation.

2.3 | Computer Simulations

We conducted individual-based simulations using R (version 4.2.2; R Core Team 2022). Three estimators of temporal F are commonly used: \hat{F}_c (Nei and Tajima 1981); \hat{F}_k (Pollak 1983); and \hat{F}_s (Jorde and Ryman 2007). As a first step, we evaluated which estimator would be most suitable for use in our study. We used a Wright-Fisher model to simulate allele frequency change across many replicate 5-generation time series of constant or variable N_e ranging from 50 to 500. The evaluation criterion we used was root-mean-squared-error (RMSE), which reflects both bias and precision and is computed as the square root of the mean squared difference between the estimate and the true parameter. For this purpose, true F was taken to be $1/(2N_e)$ for the respective generation. Consistently, \hat{F}_k had the smallest RMSE, smaller even than was found for any of the three pairwise means of the estimators, or for the overall mean of all three estimators. Accordingly, all remaining evaluations used Pollak’s estimator. For diallelic loci, temporal F is the same regardless which allele is used, and \hat{F}_k is computed as

$$\hat{F}_k = \frac{(P_1 - P_2)^2}{P(1 - \bar{P})}$$

where P_1 and P_2 are frequencies of the focal allele at times 1 and 2 and \bar{P} is the mean of P_1 and P_2 . Mean \hat{F}_k is computed as a mean across loci, with monomorphic loci (with $\bar{P} = 1$ and $1 - \bar{P} = 0$) being excluded. In what follows, we drop the subscript k as all \hat{F} apply to Pollak’s estimator.

Subsequent analyses all simulated a series of 7 consecutive generations of Wright-Fisher reproduction, with separate (Type II) samples of progeny taken for genetic analysis from the parents in each generation. The 5 middle generations formed the Core analyses, as these generations all had at least one generation of sampling before and one generation after to draw on for additional information regarding genetic drift. Generations 1 and 7 were Edge generations, as their estimates could be improved by considering either one generation before or one generation after, but not both.

The general approach to obtaining an adjusted estimate of \hat{F} and its variance was iterative. For focal generation X , an adjusted estimator \hat{F}_X^{Adj} was obtained as illustrated in Equation (10). This was repeated for each generation in the time series. At this point, the whole process was repeated, as we now had improved estimates of \hat{F}^{Adj} for each individual generation to use in Equation (9). Updating the multigeneration \hat{F} values (the off-diagonal elements in Table 2) as well as the single-generation \hat{F} values consistently reduced the standard deviation of the adjusted estimates (Figure S2).

The next sets of simulations were designed to answer these remaining questions:

1. How many generations away from the focal generation provide useful information that improves the estimates?
2. What number of iterations produces optimal results?
3. What is the optimal weighting scheme for multiple estimators of \hat{F} for the same generation?

For these simulations, the evaluation criterion was Ratio = $\sigma_{\hat{F}^{\text{Adj}}} / \sigma_{\hat{F}}$, which reflects the proportional reduction in the

TABLE 2 | Triangular matrix of temporal estimators that can be computed, given the experimental design in Figure 1, which involves sampling in 4 consecutive generations. \hat{F} values along the diagonal ($\hat{F}_{1,2}$, $\hat{F}_{2,3}$ and $\hat{F}_{3,4}$) can be used to estimate N_e separately in generations 2, 3 and 4, respectively. The off-diagonal elements provide estimates of harmonic mean N_e across multiple generations.

Generation of sample 1	Generation of sample 2		
	2	3	4
1	$\hat{F}_{1,2}$	$\hat{F}_{1,3}$	$\hat{F}_{1,4}$
2		$\hat{F}_{2,3}$	$\hat{F}_{2,4}$
3			$\hat{F}_{3,4}$

standard deviation of \hat{F} . For each simulation scenario we picked a single Target N_e (300 or 600) and modelled variation above and below that target. True N_e in generations 1 and 7 were set to Target N_e . Within the 5 Core generations, three were set to Target N_e , one was set to $2 \times \text{Target } N_e$, and one was set to Target $N_e/2$, with the order being randomly scrambled each iteration. This scheme modelled a population whose effective size varied 4-fold over a short time span of 5 generations.

Tackling Question 1 first, we picked generation 4 in the middle of the time series as the focal generation. For each replicate, we computed the single generation estimate $\hat{F}_{3,4}$, as well as the full suite of multigeneration estimates that included generation 4. From the latter, we constructed additional estimates of $\hat{F}_{3,4}$ ($\hat{F}_{3,4(A)}$, $\hat{F}_{3,4(B)}$, etc.) using the approach outlined in Equation (9) and subsequent material. Across all replicates, we computed $\sigma_{\hat{F}}^2$ for each estimator and constructed a vector of weights (W) that were inverses of $\sigma_{\hat{F}}^2$ —hence theoretically optimal weighting, assuming the estimators are independent. Finally, using these weights, for every replicate we computed adjusted $\hat{F}_{2,3}^{\text{Adj}}$ values for composite estimates that included 1, 2 or 3 additional generations of data, and across all replicates, we computed $\sigma_{\hat{F}_{2,3}^{\text{Adj}}}^2$ for all these different adjusted estimates. Inspection of the results showed that for most scenarios, adding information from one additional generation before and after the focal generation reduced $\sigma_{\hat{F}_{2,3}^{\text{Adj}}}^2$, but that including data for generations farther removed did not (Figure S3). The exceptions were for $S=25$ with Target $N_e = 300$ and $S=50$ for Target $N_e = 600$; in these two scenarios, using two additional generations improved precision, but only slightly. Accordingly, the remainder of our analyses concentrated on the focal generation, plus one generation on either side.

Subsequent simulations using one additional generation of data before and after the focal generation showed that 3 iterations of updating \hat{F} produced optimal improvements to precision (Figure S4). The only exception to this pattern was for the largest sample size ($S=200$), for which 2 iterations were slightly better than 3. Accordingly, the remainder of our analyses used 3 iterations to update \hat{F} .

Finally, we evaluated how $\sigma_{\hat{F}^{\text{Adj}}}$ varied with a range of weighting schemes of the form $\mathbf{W}=[W_1, W_2, W_3]$, where W_2 is the weight for the single-generation \hat{F} and W_1 and W_3 are weights for estimators that include one additional generation of drift before and after the focal generation, respectively. We considered symmetric scenarios with $W_1=W_3=(1-W_2)/2$, so that $\Sigma \mathbf{W}=1$ and the weighting scheme could be summarised with a single parameter (W_2). Results (Figure 2) showed that maximum reduction in $\sigma_{\hat{F}^{\text{Adj}}}$ occurred with $W_2=0.5$ and $W_1=W_3=0.25$, the only exception being that $\mathbf{W}=[0.2, 0.6, 0.2]$ was very slightly better for the scenario with $S=200$ and Target $N_e = 300$. For the Edge generation estimates, which were based on only two estimators—the single-generation \hat{F} , and another that included data from one additional generation either before or after—optimal weighting schemes were close to $\mathbf{W}=[0.4, 0.6]$ or $\mathbf{W}=[0.6, 0.4]$, with the larger weight applying to the single-generation \hat{F} .

With these preliminary evaluations out of the way, we were ready to evaluate performance of *MAXTEMP* for a range of

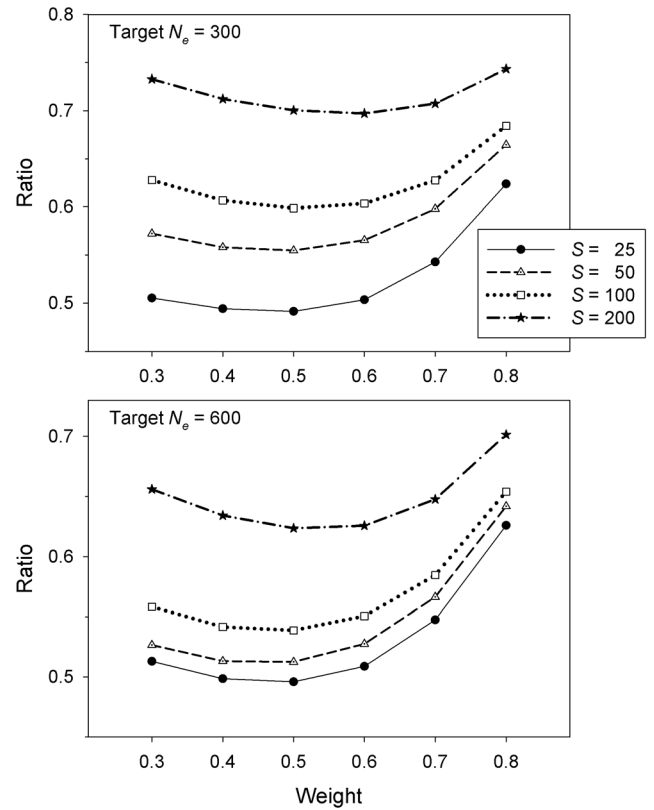


FIGURE 2 | Summary of simulation results evaluating the optimal weighting scheme for updating \hat{F} . Ratio is the ratio of $\text{sd } \hat{F}^{\text{Adj}}$ to $\text{sd } \hat{F}$ for the original single-generation estimate. Results are for \hat{F}^{Adj} values that include information from one extra generation both before and after the focal generation. Weight (X-axis) is the weight given to the single-generation \hat{F} , and each of the other two multigeneration estimates are assigned a weight of $(1 - \text{Weight})/2$. Results are shown for 4 different numbers of offspring sampled ($S=25-200$) and Target N_e of 300 (top) and 600 (bottom), with \hat{F} averaged over 600 diallelic loci.

parameter values for key covariates of interest to researchers, which (in addition to N_e) include sample sizes of individuals (S) and loci (L , assumed here to be diallelic SNPs). The simulation framework was similar, with the number of generations, number of iterations, and weighting scheme fixed as described above. We simulated all possible combinations of the following covariate values: $S=[25, 50, 100, 200]$; $L=[100, 600, 2000]$; $N_e=[50, 75, 100, 150, 200, 300, 400, 500, 650, 800, 1000]$. For each of the 132 parameter combinations, we ran 1000 replicates and recorded data for all generations with true $N_e = \text{Target } N_e$.

Results of these simulations (Table S2) were used to model the relationship between the response variable ($\sigma_{\hat{F}^{\text{Adj}}} / \sigma_{\hat{F}}$ = 'Ratio') and the 3 covariates. Ratios for 3 of the 5 Core generations that equaled the Target N_e were treated as replicates in the modelling. Based on exploratory plots of the simulated data, we regressed log-transformed Ratio on log-transformed covariates N_e , S and L , including a quadratic term for log-transformed N_e :

$$Y_{ijkl} = \beta_0 + \beta_1 \ln(N_{e(i)}) + \beta_2 \ln(N_{e(i)})^2 + \beta_3 \ln(S_j) + \beta_4 \ln(L_k) + \varepsilon_{ijkl}$$

where Y_{ijkl} = natural log-transformed Ratio $\sigma_{\hat{F}^{\text{Adj}}} / \sigma_{\hat{F}}$ and $\varepsilon_{ijkl} \sim N(0, \sigma_\varepsilon^2)$. Log-transformations of the response variable and

covariates improved model assumptions of linearity, constant variance and normality.

We also ran simulations to illustrate the second capability of *MAXTEMP*: compensating for missing data. The LD method (Waples and Do 2008) was used to generate single-sample estimates of N_e . We used a Target N_e of 100 and ran 6 generations of burn-in to allow levels of LD to reach a dynamic equilibrium, followed by 2 more generations of drift (Figure 3). Multilocus genotypes were tracked for 1000 unlinked loci. In the Constant scenario, Generations 7 and 8 also simulated Target N_e , whereas in the Change scenarios, Generations 7 and 8 simulated either $N_e = 125$ and 75, respectively (Change/Up), or $N_e = 75$ and 125 (Change/Down). Samples of $S = 25, 50$ or 100 offspring were taken of progeny from generations 6 and 8 but not 7. The samples from generations 6 and 8 were used to compute $\hat{F}_{6,8}$, which when adjusted for sample size and used in Equation (5) with $t = 2$, was expected to produce $\hat{N}_{e7-8} = 100$ for the Constant scenario and $\hat{N}_{e7-8} = 93.75$ (the harmonic mean of 75 and 125) for the two Change scenarios. The expected value of estimated N_e for the LD sample from generation 8 was either 75 (Change/Up) or 125 (Change/Down), while the true (mystery) N_e for unsampled generation 7 was the reverse (125 for Change/Up and 75 for Change/Down). Following Equation (2), Mystery N_e for generation 7 was estimated as

$$\hat{N}_{e7} = \frac{1}{\frac{2}{\hat{N}_{e7-8}} - 1/\hat{N}_{e8LD}}$$

where \hat{N}_{e8LD} was the LD estimate and \hat{N}_{e7-8} was derived from $\hat{F}_{6,8}$.

2.4 | Empirical Example

We illustrate practical application of the new method using empirical data for a long-term genetic monitoring program for delta smelt, *Hypomesus transpacificus*, which is endemic to the San Francisco Estuary in California, USA. The delta smelt is a useful system for testing *MAXTEMP* because the species has a nearly annual life

cycle (Moyle et al. 1992) and each year effectively represents a single genetic population (Fisch et al. 2011). In addition, due to the complexity of and human impact on the San Francisco Estuary, a variety of systematic, long-term monitoring surveys have been conducted in the area, some for decades (Tempel et al. 2021). Methodologies for these surveys are largely unchanged over time, and survey crews record all species captured and take samples as needed, providing a rich archive of DNA from delta smelt that can be used for genetic analysis. In particular, estimating N_e has been of great interest to managers, especially as the wild population nears extinction and a program of experimental releases from the conservation hatchery has begun (USFWS 2020, 2022).

For this example, we used previously generated RAD-sequencing data for wild delta smelt obtained in state and federal trawls from 2013 to 2020 (SRA BioProject PRJNA1112857; <https://www.ncbi.nlm.nih.gov/sra>). For more details on sample sourcing, sequencing and data processing, see [Supporting Information](#). We used 1114 SNP loci for samples of 50 individuals each for the years 2013–2020, inclusive (Table S3). Samples were assigned to cohorts for N_e estimation based on survey timing and gear used. Delta smelt typically spawn in February–May (Moyle et al. 1992). For example, samples assigned to the 2013 cohort are offspring from spawning events in 2013. The full matrix of temporal \hat{F} values using Pollak’s estimator were computed in NeEstimator (Do et al. 2014). For each year (=each generation), we used *MAXTEMP* to compute adjusted \hat{F}^{Adj} values, which were used to provide improved estimates of N_e each generation. Based on predictions from our linear model, given the specific covariate values applicable to each year, we generated improved CIs for each of these N_e^{Adj} point estimates.

3 | Results

3.1 | Adjusted Temporal Estimates

The best-fit model used natural log to transform Ratio = $\sigma_{\hat{F}^{Adj}} / \sigma_{\hat{F}}$ = the ratio of the standard deviation of the adjusted \hat{F} to the standard deviation of the initial, single-generation \hat{F} and all

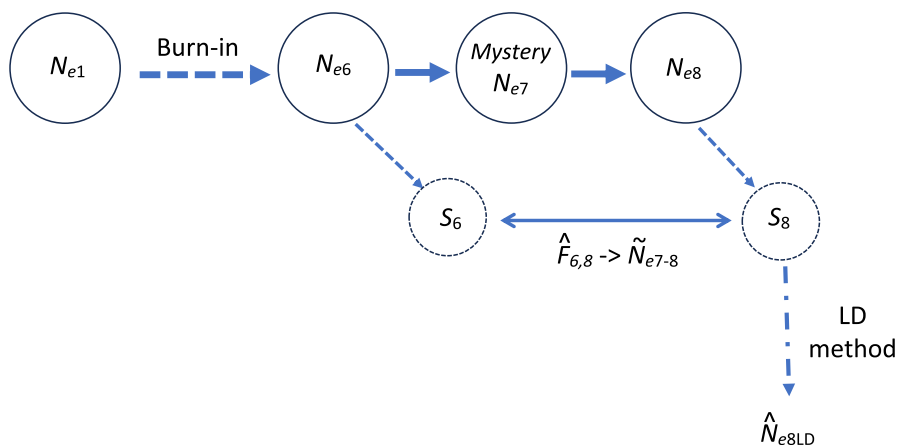


FIGURE 3 | Schematic of the scenario used to model compensating for a missing sample. Samples of offspring are taken from generations 6 and 8 but not 7. Temporal F comparing allele frequencies in the two samples ($\hat{F}_{6,8}$) provides an estimate of the harmonic mean variance N_e in generations 7 and 8. If a single-sample method (in this case the LD method) can be used to estimate N_{e8} , the ‘mystery’ N_e in unsampled generation 7 can be estimated using a variation of Equation (2).

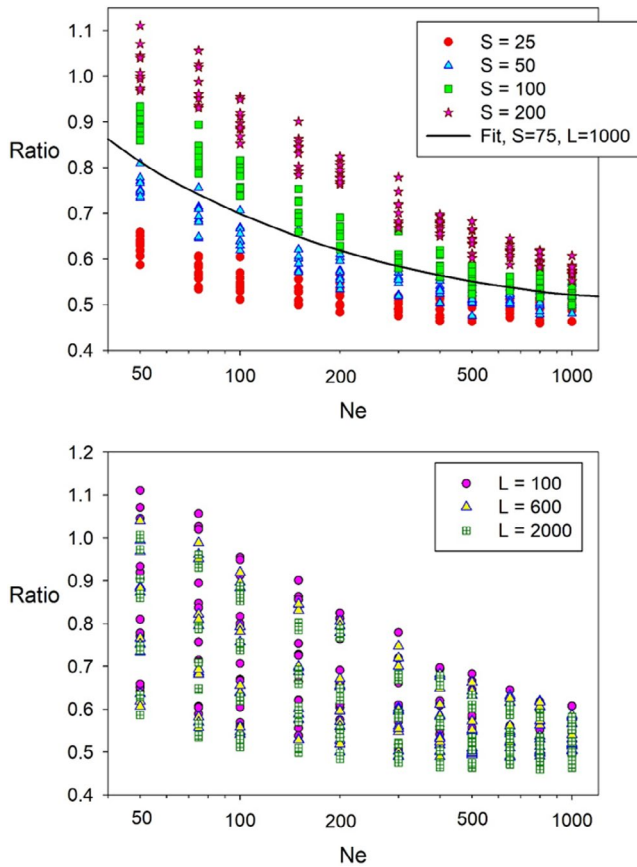


FIGURE 4 | Simulation results quantifying increases in precision for \hat{F} . Symbols are colour coded according to sample size (S , top) or number of loci (L , bottom). Ratio = $\sigma_{\hat{F}^{Adj}} / \sigma_{\hat{F}}$ is the ratio of the standard deviation of \hat{F}^{Adj} to the standard deviation of the initial, single-generation \hat{F} . The black curve in the top panel is the predicted $\sigma_{\hat{F}^{Adj}} / \sigma_{\hat{F}}$ assuming $S=75$ and $L=1000$. Note the log scale on the X -axis. These results are for estimates that can leverage information from one additional generation of genetic drift both before and after the focal generation. See Figure S5 for results for estimates that can use information from one additional generation of genetic drift either before or after the focal generation, but not both.

covariates. All covariates were highly significant, and the model explained 92.1% of the variance in the data. The estimated prediction equation for Ratio was

$$\widehat{\text{Ratio}} = \exp. \left(0.627 - 0.477 * \ln(N_e) + 0.0304 * \ln(N_e)^2 + 0.169 * \ln(S) - 0.0236 * \ln(L) \right) \quad (12)$$

The response variable, $\widehat{\text{Ratio}}$, allows one to estimate the percent reduction in $\sigma_{\hat{F}}$ that is possible from utilising more of the data:

$$\text{Estimated \% reduction in } \sigma_{\hat{F}} = 100 * \left(1 - \widehat{\text{Ratio}} \right)$$

As shown in Figure 4, for $N_e < 100$ the percent reduction in $\sigma_{\hat{F}^{Adj}}$ compared to $\sigma_{\hat{F}}$ is only about 10%–30%, depending on values of the other covariates, but for moderate to large N_e $\sigma_{\hat{F}^{Adj}}$ can be reduced by nearly half.

Hypothetical data in Figure 5 illustrate the dramatic effect the reduced variance associated with \hat{F}^{Adj} can have for CIs around \hat{N}_e . Assuming accurate point estimates for true N_e from 50 to

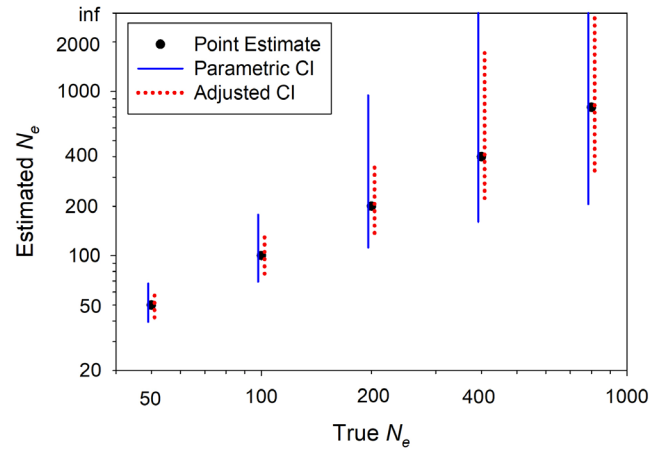


FIGURE 5 | Point estimates and confidence intervals for hypothetical scenarios with $S=50$, $L=1000$ and true $N_e=50$ –800. 95% CIs are arranged around point estimates (black circles) assumed to agree with true N_e . CIs on the left side of point estimates (solid blue lines) are based on parametric expectations for $\sigma_{\hat{F}}$, and CIs on the right (red dotted lines) are based on modelled $\sigma_{\hat{F}^{Adj}}$. The upper bound for the parametric CI for true $N_e=400$ and both upper bounds for true $N_e=800$ extend to infinity ('inf'). Note the log scale on both axes.

800, the parametric CIs using only the single-generation \hat{F} are shown on the left and the adjusted CIs that take into consideration one additional generation of data on either side of the focal generation are shown on the right. Relative improvement of the CIs increases with N_e and is particularly noticeable for $N_e=200$. For $N_e=400$, the adjusted CI has an upper bound of just under 2000, while the upper bound to the parametric CI is infinity. For $N_e=800$, the upper bound of both CIs is infinity, but the adjusted CI has a considerably higher lower bound (329 vs. 205).

These results assume that at least one generation of additional data are available both before and after the focal generation. When the focal generation is at either the start or end of a genetic monitoring program, the potential benefits are reduced. In that scenario, the best-fit model was

with an adjusted R^2 of =0.849. With moderate to large N_e , the maximum percent reduction in $\sigma_{\hat{F}^{Adj}}$ is about 30% rather than ~50% when the focal generation is in the middle of the time series (Figure S5).

3.1.1 | Reduction of Negative Estimates

In moment-based estimators like the standard temporal method, the expected contribution from sampling error is subtracted from the raw F , producing F' which in theory only reflects genetic drift (see Equation 5). In practice, both F and F' are random variables, and F' can sometimes be negative just by chance, particularly

when true N_e is large or only modest amounts of data are available. Negative F' translates to a negative estimate of N_e , which is interpreted as $\hat{N}_e = \infty$ (no evidence for genetic drift).

$$\widehat{Ratio} = \exp. \left(0.2688 - 0.2042 \times \ln(N_e) + 0.01079 * \ln(N_e)^2 + 0.09408 \times \ln(S) - 0.0128 * \ln(L) \right) \quad (13)$$

MaxTemp does not eliminate negative F' estimates but, because it lowers the variance of \hat{F} , it reduces their incidence. The absolute reduction in the fraction of negative F' values was generally greater when the initial (raw) fraction was higher, but proportionally the reductions were greatest when the raw fraction was low (Figure S6). For example, with $N_e = 50$, $S = 25$ and $L = 600$, the fraction of negative F' was 0.025 in the raw data and 0.001 after adjusting \hat{F} —a reduction of 97%, although the absolute reduction was small (Table 3). In the scenario with $N_e = 100$, $S = 100$ and $L = 600$, the drop in the fraction of negative F' was larger in absolute terms (0.225 in the raw data and 0.073 after adjustment), but this represented only a two-thirds drop in the incidence of negative F' . Patterns shown in Table 3 can be predicted theoretically, which allows one to predict the fractions of negative F' values as a function of N_e , S and L (see [Supporting Information](#), Table S4).

3.2 | Empirical Example

Increased precision for temporal estimates provided by *MAXTEMP* is nicely illustrated in the delta smelt example (Figure 6). After adjustment, individual Wald CIs were narrower for 6 of the 7 generations. The only exception was 2014, for which only one extra datapoint was available to improve the estimate; the adjusted point estimate also increased considerably, which (all else being equal) leads to wider CIs. The direct benefits of *MAXTEMP* are most easily seen for generations 2015 and 2018. In 2015, the adjusted point estimate is slightly higher, but nevertheless the upper bound of the CI is finite, whereas it extended to infinity in the original data. In 2018, the initial and adjusted point estimates are nearly the same, but the width of the adjusted 95% CI is much narrower (148–480 vs. 110–1424). The new results also have direct conservation relevance. The lower bounds of all adjusted CIs were higher, which is good news from a conservation perspective. For the initial estimates, lower bounds of 6 of the 7 initial CIs dipped below 200, and the 7th (for 2019) almost did. After applying *MAXTEMP*, only 2 adjusted CIs dropped this low.

3.3 | Compensating for a Missing Sample

3.3.1 | Bias

In the Constant N_e scenario of the missing sample method (Figure 7, top), the temporal-method estimate for generations 7–8 was very close to the expected value of 100, whereas the LD estimate for generation 8 was on average about 3%–5% too high. With these two estimates as inputs, the estimate of Mystery N_e in the unsampled generation 7 was a bit low on average, from about 3% for $S = 100$ to about 10% for $S = 25$.

In the population-change scenarios (Figure 7, middle and bottom), the temporal-method estimate was very close to the expected 94, which is the harmonic mean of 75 and 125. In the

Change/Down scenario, where N_e first dropped to 75 in generation 7 before climbing to 125 in generation 8, the LD estimate for generation 8 was slightly low for all sample sizes, and the estimate for the unsampled generation 7 was essentially unbiased (Figure 7, middle). With the reverse pattern (Change/Up, where N_e rose to 125 before declining to 75), the LD method overestimated N_{e8} by about 10%–15%, and the estimate of Mystery N_{e7} was correspondingly biased downwards (Figure 7, bottom).

3.3.2 | Precision

Consistently, empirical CIs to \hat{N}_e were tightest for generation 8 using the LD method and widest for the unsampled generation 7. As expected, regardless of the method, CIs were narrower for larger sample sizes. Reduced precision from small samples was particularly noticeable for Mystery N_{e7} . With $S = 25$, upper CI bounds were > 1000 for all three scenarios and infinitely large in two of the three, whereas with $S = 100$ the upper CI bounds were all < 215 .

4 | Discussion

Two previous studies using the moment-based temporal method have incorporated information from multiple samples to estimate N_e : Pollak (1983) for the standard temporal method that assumes discrete generations, and Jorde and Ryman (1995) for a modification that accounts for age structure and overlapping generations. However, both of these methods are designed to estimate a single N_e , which is assumed to be constant over time. For the unusual life history of Pacific salmon (which are semelparous but have variable age at maturity), Waples, Masuda, and Pella (2007) developed a version of the temporal method that uses multiple samples to estimate the effective number of breeders (N_b) in individual years, which represent a fraction of a generation. In contrast to these approaches, *MAXTEMP* is designed to maximise information regarding generational N_e in specific generations.

A variety of authors have proposed variations of the temporal method that, instead of computing the standardised variance F and relating it to N_e , focus on maximum likelihood (ML) analysis of the sampled allele frequencies. The ML estimate of effective size is the value of N_e that has the highest probability of producing the observed allele frequencies (Edwards 1992). The likelihood is calculated as a sum over all the unobserved population allele frequencies, which means that additional samples in time provide more information about the unobserved population allele frequencies at every time period. In this respect, the ML methods already leverage information contained in multiple samples. In theory, this information

TABLE 3 | Fraction of simulated datasets with negative \hat{F}' values (which lead to infinite estimates of N_e) for raw data ('Original') and after adjusting with MaxTemp ('Adjusted'). S =number of offspring sampled; L =number of loci used to compute \hat{F}' . The last column shows the fraction of raw negative \hat{F}' values that remain negative after adjustment. A: \hat{F}'^{Adj} is based on one extra generation both before and after the focal generation. B: \hat{F}'^{Adj} is based on one extra generation either before or after the focal generation, but not both.

N_e	S	L	Original	Adjusted	Adj/Orig
A					
Fraction $\hat{F}' < 0$					
50	25	600	0.025	0.001	0.027
		2000	0.010	0.001	0.065
50	100	600	0.000	0.000	—
		2000	0.000	0.000	—
200	25	600	0.187	0.059	0.317
		2000	0.123	0.014	0.114
200	100	600	0.001	0.000	0.000
		2000	0.000	0.000	—
1000	25	600	0.367	0.316	0.863
		2000	0.284	0.209	0.735
1000	100	600	0.225	0.073	0.323
		2000	0.093	0.007	0.075
B					
Fraction $\hat{F}' < 0$					
50	25	600	0.017	0.011	0.618
		2000	0.008	0.006	0.721
50	100	600	0.000	0.000	—
		2000	0.000	0.000	—
200	25	600	0.176	0.127	0.721
		2000	0.109	0.061	0.555
200	100	600	0.002	0.001	0.333
		2000	0.000	0.000	—
1000	25	600	0.359	0.363	1.011
		2000	0.287	0.281	0.979
1000	100	600	0.217	0.140	0.644
		2000	0.091	0.044	0.481

could be used to estimate N_e in separate time periods, which is the goal of *MaxTemp*. In practice, doing that has not been the focus of authors of ML versions of the temporal method. The initial likelihood papers (Williamson and Slatkin 1999; Anderson, Williamson, and Thompson 2000) allowed for multiple samples but assumed that N_e is constant. Variations based on the coalescent (Berthier et al. 2002; Anderson 2005) considered only a pair of samples. The basic model for Wang's (2001) pseudo-likelihood method also assumes a pair of temporal samples. However, Wang outlined a 3-sample version that allows estimation of N_e in 2 different time periods, as well as an exponential-growth model that assumes that N_e grows or shrinks each generation at a constant rate (an idea

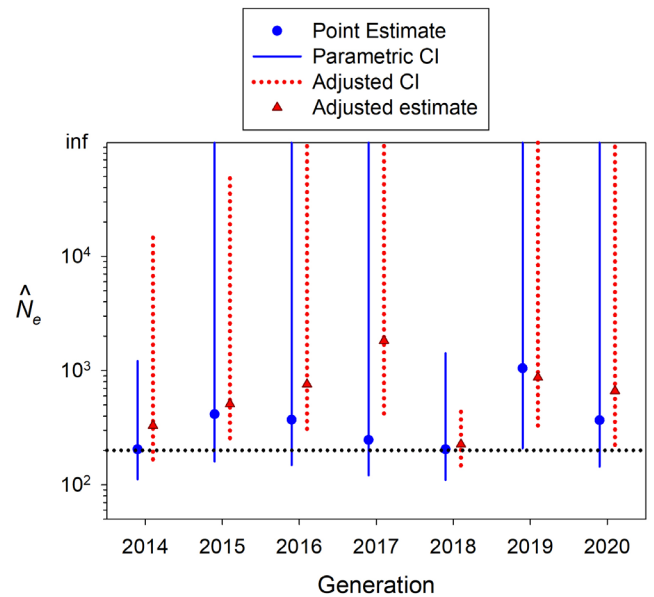


FIGURE 6 | Point estimates of N_e and 95% confidence intervals for 7 generations of delta smelt data. Initial point estimates and CIs on the left (blue circles and solid blue lines) are based on initial \hat{N}_e and parametric expectations for $\sigma_{\hat{F}'}$, and adjusted point estimates and CIs on the right (red triangles and red dotted lines) are based on \hat{N}_e^{Adj} and modelled $\sigma_{\hat{F}'^{\text{Adj}}}$. Upper bounds of some CIs extend to infinity (inf). Note the log scale on the Y axis.

introduced by Felsenstein 1971). Finally, the ML version proposed by Hui and Burt (2015) uses a more efficient ML algorithm to alleviate some of the computational limitations of earlier likelihood approaches. Their base model assumed two samples and a fixed N_e . An alternative model allows for 3 or more samples, and Hui and Burt noted that it is possible to modify their model to allow N_e to vary, but this idea is not examined in any detail. The Hui and Bert software package NB has no provision for estimating different N_e in different time intervals. So, although some previous authors of ML papers have fitted multiple N_e values with 3 or more temporal samples, these features have not been incorporated into their respective software. Directly comparing performance of *MaxTemp* and the ML methods for this specific purpose could be a topic for future research.

Major themes related to practical implementation of the new method are summarised below.

4.1 | Adjusted Temporal Estimates

By leveraging additional information about genetic drift contained in multigeneration estimates of temporal F that are routinely generated as part of a genetic monitoring program, the standard deviation of \hat{F}' applicable to individual generations can be reduced by up to about 50% for generations that are interior to the series and by up to about 30% for the first and last generations. Proportional increases in precision are greatest when true N_e is large and sample sizes of individuals are small. This is good news for researchers, as those are the scenarios

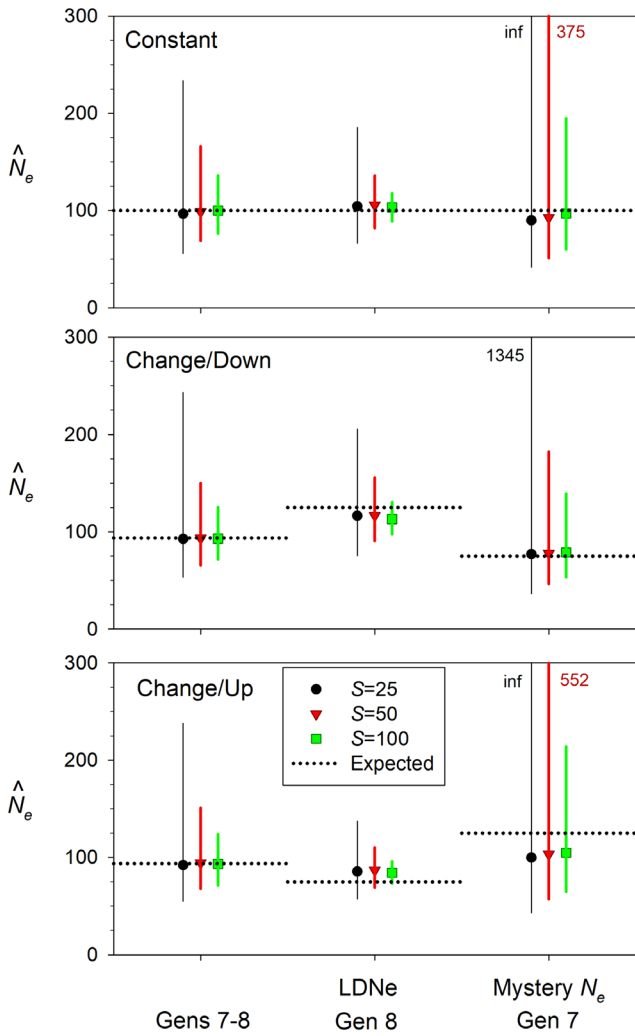


FIGURE 7 | Results of performance evaluation of the method to estimate N_e in an unsampled generation. See Figure 3 for a schematic representation of the experimental design. True N_e in generations 6–8 is either [100/100/100] (top panel), [100/75/125] (middle) or [100/125/75] (bottom). Samples of $S=25$, 50 or 100 progeny are taken from generations 6 and 8, but not 7, and genotyped for 1000 diallelic loci; symbols show resulting harmonic mean \hat{N}_e across 1000 replicates. Samples from generations 6 and 8 were used to compute $\hat{F}_{6,8}$ and this was used to estimate the harmonic mean N_e in generations 7 and 8 (left). The sample from generation 8 was used to estimate N_{e8} with the LD method (middle), and this information was used in conjunction with information from the temporal samples to estimate Mystery N_e in the unsampled generation 7 (right). Horizontal dotted lines represent true N_e for generations 7 and 8 and harmonic mean true N_e across both generations. Vertical coloured lines for each sample size show empirical 95% confidence intervals for \hat{N}_e ; coloured numbers indicate upper CI bounds that exceeded 300.

for which it is most difficult to obtain robust estimates of effective size. All else being equal, a reduction to $\sigma_{\hat{F}}$ also reduces the size of confidence intervals around \hat{N}_e . For example, with samples sizes of $S=50$ individuals scored for $L=1000$ SNPs and true $N_e=200$, expected 95% confidence intervals to \hat{N}_e would be 112–947 (range=835) based on \hat{F} spanning a single generation but would drop to 137–367 (range=230) after

integrating information from two other multigeneration estimates (Figure 5).

In addition to narrower CIs, a second benefit to the new method is to stabilise the point estimates of \hat{N}_e , such that on average they span a narrower range. In any given empirical dataset, this general pattern might be manifest in a variety of ways. In the delta smelt example (Figure 6), all but one of the adjusted estimates were higher than the raw \hat{N}_e . This result, combined with tighter adjusted CIs for every generation except 2014, presents a slightly more favourable picture of the species' conservation status than did the raw data.

It should be pointed out that, regardless whether raw or adjusted estimates are used, there is a positive correlation between \hat{N}_e and CI width, such that (for fixed sample sizes of individuals and loci) CIs are wider when \hat{N}_e is larger. For example, for the standard temporal method using data shown in Figure 5, parametric 95% CIs are 40–68 for true $N_e=50$ and 160 to infinity for true $N_e=400$. This means that if adjusted point estimates of N_e increase using *MAXTEMP*, adjusted CIs might also be wider, even if $\sigma_{\hat{F}}$ is reduced. This effect is seen for 2014 in the delta smelt example (Figure 6): the point estimate for that year increased by over 50% (from $\hat{N}_e=204$ –330) and the lower bound of the CI rose sharply as well (from 111 to 167), but the upper bound for the adjusted CI rose from 1208 to almost 15,000.

If $\sigma_{\hat{F}}$ for individual generations continued to decline as more and more extra generations of data were used, long-term genetic monitoring programs could be particularly effective in increasing precision of temporal estimates of N_e . That, however, proved not to be the case, as using more than one additional generation on either side of the focal generation did not continue to increase precision. We expect that this result reflects the non-independence of estimators using overlapping sets of the same generations. The default inverse-variance weighting scheme is optimal only if the elements being weighted are independent, but that is not the case here. We are interested in the variance of the adjusted estimator \hat{F}^{Adj} , which is the weighted sum of two or more terms (call them A and B). Standard theory tells us that $\text{var}(A+B) = \text{var}(A) + \text{var}(B) + 2\text{cov}(A, B)$. In our experimental design, the pairwise covariance terms are positive and therefore affect the optimal weighting scheme. Consider, for example, a design with samples of progeny taken from 3 consecutive generations, 1–3, leading to the single-generation estimator $\hat{F}_{1,2}$ and the multi-generation estimator $\hat{F}_{1,3}$. These two estimators can be used as described above to derive the adjusted estimator $\hat{F}_{1,2}^{\text{Adj}}$, which then can be used to estimate N_e in generation 2. Now consider a fourth sample, from generation 4, and the new estimator $\hat{F}_{1,4}$, which contains information about genetic drift transitioning from generation 1 to 2, from 2 to 3 and from 3 to 4. However, $\hat{F}_{1,2}^{\text{Adj}}$ already includes 2 separate estimates of genetic drift for generations 1–2 and one estimate for generations 2–3. Furthermore, $\hat{F}_{1,2}$, $\hat{F}_{1,3}$, and $\hat{F}_{1,4}$ all use the same sample from generation 1 to establish initial allele frequencies. All of these factors create a large positive covariance between the drift signal contained in $\hat{F}_{1,4}$ and information that is already reflected in $\hat{F}_{1,2}^{\text{Adj}}$. In effect, these covariance terms create a barrier that must be surmounted before a new estimator usefully can be

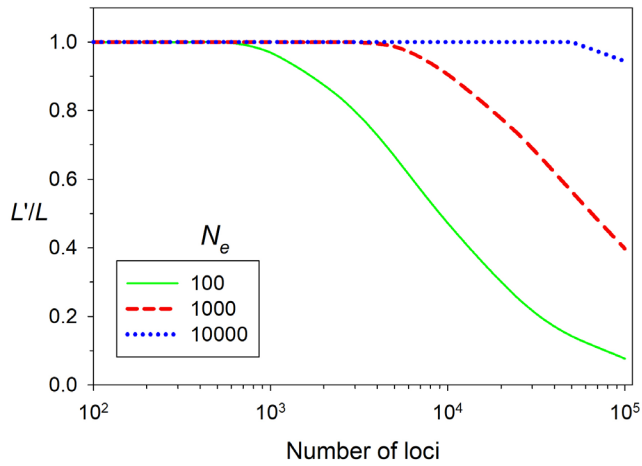


FIGURE 8 | The ratio of the effective number of diallelic loci used to estimate F_{ST} (L') to the actual number of loci (L), as a function of N_e and L . L' is defined as the number of theoretically independent loci that would produce the same variance of mean F_{ST} as the L loci actually used. Put another way, the L linked loci have the same information content as would L' totally independent loci. Plots are based on results from Waples, Waples, and Ward (2022), assuming the species has 20 chromosomes and sample sizes are 50 individuals. The distribution of temporal F is similar to that for F_{ST} so results should be applicable to analyses reported here.

added into the mix. In our preliminary modelling, positive covariances associated with any estimators extending more than one generation beyond the focal generation were large enough that attempting to include them in the analysis degraded rather than improved performance.

4.1.1 | Genomics-Scale Datasets

Analytical and simulation results presented here have assumed that the loci used are all unlinked and hence provide independent information about genetic drift. In reality, all loci have to be apportioned among a relatively small number of chromosomes (mean = 25 in vertebrates; Li et al. 2011), within which recombination is relatively rare. As more and more loci are used to estimate key genetic metrics like temporal F , precision does not increase as fast as it would if all the datapoints were independent. Waples, Waples, and Ward (2022) quantified the effects of this lack of independence (aka pseudoreplication) on precision of mean r^2 (a measure of LD) and F_{ST} . The latter has distributional properties similar to temporal F , so the Waples, Waples, and Ward (2022) results for F_{ST} are relevant here. See Hui, Brenas, and Burt (2021) for an alternative approach for evaluating lack of independence of loci for estimating temporal F . If the loci are all independent, the variance of \hat{F}_{ST} (and \hat{F}) should be inversely proportional to the number of diallelic loci, L : $\text{var}(\hat{F}_{ST}) = \sigma_{\hat{F}}^2 = 2\hat{F}_{ST}^2 / L$. In this equation, L can be considered to be the degrees of freedom associated with the estimate. Waples et al. showed that the actual (effective) degrees of freedom for F_{ST} (L') was a function of true N_e , S , L and genome size (number of chromosomes) and that using L' rather than L in the above equation accurately predicted the true variance of \hat{F}_{ST} . Unless N_e is small (< 100), L' is appreciably less than L only when more than 1000 loci are used, and if N_e is moderately large (~ 1000) there is little pseudoreplication

unless about 10^4 or more loci are used (Figure 8). Hence, lack of independence is expected to have had little effect on the analyses reported here. However, large genomics-scale datasets can have 10^4 – 10^7 loci even for non-model species, so this issue is important to consider more generally. *MAXTEMP* incorporates results from Waples, Waples, and Ward (2022) that allow it to predict L' and L'/L for temporal F , given known values of S and L and estimates of N_e and genome size.

4.2 | Compensating for a Missing Sample

Collecting samples from natural populations is often logistically challenging, time consuming and expensive, so occasional gaps are not uncommon in genetic monitoring programs. Here, we showed how a one-generation gap can be overcome, using a combination of temporal and single-sample estimators to obtain an estimate of N_e in the unsampled generation. In our model, the temporal method was used to estimate harmonic mean N_e in generations 7 and 8, and these estimates were largely unbiased, both for constant N_e and variable N_e (Figure 7, left). Accuracy of the estimates of mystery N_e in the unsampled generation 7 therefore depends on accuracy of the single-sample estimate for generation 8. With constant N_e , the LD method tends to slightly overestimate effective size, which leads to a corresponding slight underestimate of N_e in generation 7 (Figure 7, top). When N_e changes over time, the LD method is influenced to some extent by N_e in previous generations. Under the Change/Down scenario (true $N_e = 75$ in generation 7 and 125 in generation 8), the LD estimate for generation 8 was biased slightly downwards by the lower N_e in generation 7, but the estimate of mystery N_e showed essentially no bias (Figure 7, middle). Under Change/Up (true $N_e = 125/75$; Figure 7 bottom), the higher N_e in generation 7 pushed the LD estimate for generation 8 higher, and (combined with the slight overall tendency to overestimate N_e) this caused the overall LD estimate for generation 8 to be about 10%–15% too high. This in turn caused an underestimate of N_e in unsampled generation 7.

General conclusions regarding bias that emerge from our results are as follows. If N_e is constant or nearly so, or if N_e is lower in the unsampled generation, the estimate of mystery N_e will be unbiased or nearly so, with somewhat better performance for larger sample sizes. If N_e is larger in the unsampled generation, effects of prior N_e and the tendency to slightly overestimate N_e reinforce each other, with the result that mystery N_e can be underestimated to some extent. In this case, an option would be to use Wang's (2009) sibship method to provide the single-sample estimate of N_e . The incidence of siblings should not be affected by N_e in previous generations, which could lead to less bias.

It is apparent from Figure 7 that there is a cost in terms of reduced precision for having to estimate Mystery N_e without the benefit of a directly relevant sample. Fortunately, this reduced precision can be overcome to some extent by taking larger samples of individuals from other generations.

Some readers will have noticed that joint use of the temporal and LD methods to estimate Mystery N_e in an unsampled generation produces a mixture of estimates of variance N_e and inbreeding N_e , which are two related but different ways of characterising

random genetic processes. The LD method estimates inbreeding N_e and relates to effective size in the parental generation; the temporal method estimates variance N_e and relates to effective size in the offspring generation (Crow 1954). In combining results for the two methods, the key is to focus on the time period(s) to which each estimate applies (Waples 2005). With respect to the experimental design used here, the LD method sample of progeny from generation 8 estimates N_{e8} and the temporal-method samples of progeny from generations 6 and 8 estimate \tilde{N}_{e7-8} . The drift signal from generation 7 is contained within \hat{N}_{e7-8} , and information from \hat{N}_{e8LD} allows one to tease that signal out.

4.3 | Caveats and Limitations

Several caveats are in order regarding the new methods proposed here. First, like the standard temporal method and most other genetic methods for estimating N_e , *MAXTEMP* assumes that generations are discrete. Most species in nature are age-structured, and age structure introduces potential biases for the temporal method that depend, among other things, on the experimental design and the species' life history (Waples and Yokota 2007). Second, the treatment here assumes sampling is according to Plan II, where individuals comprising the initial sample are collected before reproduction and removed so that they cannot contribute genes to subsequent generations. If the initial individuals are sampled after reproduction or non-lethally before reproduction, sampling is according to Plan I and it is necessary to add a term for $1/N$ into Equations (3–5), where N is the total number of individuals subject to sampling the first sample (Nei and Tajima 1981; Waples 1989). The model could be modified to account for Plan I sampling, but that is not attempted here. Third, the standard temporal method assumes a closed population, and estimates can be biased if immigrants with different allele frequencies enter the population between sampling events. Effects of migration are generally relatively small over short time periods unless the migration rate is fairly high (Luikart et al. 2010; Gilbert and Whitlock 2015). Finally, the temporal method is based on the theoretical drift variance in frequency of neutral alleles, and estimates can be biased by strong directional or stabilising selection. Researchers might consider using one of the many software programs that attempt to identify 'outlier' loci, which show unusually high levels of allele frequency change over time. Filtering out such loci might provide a more reliable indicator of random genetic drift (Therkildsen et al. 2013).

Acknowledgements

Genetic work for delta smelt was supported through Bureau of Reclamation Grants #R10AC20089, #R15AC00030 and #R20AC00027, and State Water Contractors Agency Agreement A19-1844. Delta smelt samples were collected by the University of California, Davis Fish Culture and Conservation Laboratory, the California Department of Fish and Wildlife, and the U.S. Fish and Wildlife Service. Delta smelt genetic data were generated by Alisha Goodbla, Emily Funk, Mary Badger and Grace Auringer. The authors would like to thank Jim Hobbs and the Otolith Geochemistry and Fish Ecology Lab at UC Davis for assistance with assigning delta smelt samples to cohorts, and Eric Anderson for discussion of likelihood methods for estimating N_e . Tin-Yu J. Hui and Julian Wittische provided valuable comments on an earlier draft. The scientific results and conclusions, as well as any views or opinions

expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Genetic data and associated metadata used for delta smelt data are deposited in the SRA (BioProject PRJNA1112857; <https://www.ncbi.nlm.nih.gov/sra>). The R code that implements *MAXTEMP* is available on Zenodo (<https://doi.org/10.5281/zenodo.14213748>) and can be cited as Waples et al. (2024). R code to conduct the simulations and analyses in this paper is also posted at <https://doi.org/10.5281/zenodo.14219013>.

References

- Anderson, E. C. 2005. "An Efficient Monte Carlo Method for Estimating n_e From Temporally Spaced Samples Using a Coalescent-Based Likelihood." *Genetics* 170, no. 2: 955–967.
- Anderson, E. C., E. G. Williamson, and E. A. Thompson. 2000. "Monte Carlo Evaluation of the Likelihood for n_e From Temporally Spaced Samples." *Genetics* 156: 2109–2118.
- Berthier, P., M. A. Beaumont, J. M. Cornuet, and G. Luikart. 2002. "Likelihood-Based Estimation of the Effective Population Size Using Temporal Changes in Allele Frequencies: A Genealogical Approach." *Genetics* 160: 741–751.
- Charlesworth, B. 2009. "Effective Population Size and Patterns of Molecular Evolution and Variation." *Nature Reviews Genetics* 10, no. 3: 195–205.
- Clarke, S. H., E. R. Lawrence, J. M. Matte, et al. 2024. "Global Assessment of Effective Population Sizes: Consistent Taxonomic Differences in Meeting the 50/500 Rule." *Molecular Ecology* 33, no. 11: e17353.
- Crow, J. F. 1954. "Breeding Structure of Populations. II. Effective Population Number." In *Statistics and Mathematics in Biology*, edited by O. Kempthorne, T. Bancroft, J. Gowen, and J. Lush, 543–556. Ames, Iowa USA: Iowa State University Press.
- De Barba, M., L. P. Waits, E. O. Garton, et al. 2010. "The Power of Genetic Monitoring for Studying Demography, Ecology and Genetics of a Reintroduced Brown Bear Population." *Molecular Ecology* 19, no. 18: 3938–3951.
- Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillet, and J. R. Ovenden. 2014. "NeEstimator V2: Re-Implementation of Software for the Estimation of Contemporary Effective Population Size (N_e) From Genetic Data." *Molecular Ecology Resources* 14: 209–214.
- Edwards, A. W. F. 1992. *Likelihood*. Expanded ed. Baltimore: Johns Hopkins University Press.
- Felsenstein, J. 1971. "Inbreeding and Variance Effective Numbers in Populations With Overlapping Generations." *Genetics* 68: 581–597.
- Fisch, K. M., J. M. Henderson, R. S. Burton, and B. May. 2011. "Population Genetics and Conservation Implications for the Endangered Delta Smelt in the San Francisco Bay-Delta." *Conservation Genetics* 12, no. 6: 1421–1434.
- Foote, A. D., P. F. Thomsen, S. Sveegaard, et al. 2012. "Investigating the Potential Use of Environmental DNA (eDNA) for Genetic Monitoring of Marine Mammals." *PLoS One* 7, no. 8: e41781.
- Fussi, B., M. Westergren, F. Aravanopoulos, et al. 2016. "Forest Genetic Monitoring: An Overview of Concepts and Definitions." *Environmental Monitoring and Assessment* 188: 1–12.

- Gilbert, K. J., and M. C. Whitlock. 2015. "Evaluating Methods for Estimating Local Effective Population Size With and Without Migration." *Evolution* 69: 2154–2166.
- Hare, M., L. Nunney, M. K. Schwartz, et al. 2011. "Understanding and Estimating Effective Population Size for Practical Application in Marine Conservation and Management." *Conservation Biology* 25: 438–449.
- Hill, W. G. 1981. "Estimation of Effective Population Size From Data on Linkage Disequilibrium." *Genetics Research* 38, no. 3: 209–216.
- Hoban, S., F. I. Archer, L. D. Bertola, et al. 2022. "Global Genetic Diversity Status and Trends: Towards a Suite of Essential Biodiversity Variables (EBVs) for Genetic Composition." *Biological Reviews* 97, no. 4: 1511–1538.
- Hui, T. Y. J., J. H. Brenas, and A. Burt. 2021. "Contemporary ne Estimation Using Temporally Spaced Data With Linked Loci." *Molecular Ecology Resources* 21, no. 7: 2221–2230.
- Hui, T. Y. J., and A. Burt. 2015. "Estimating Effective Population Size From Temporally Spaced Samples With a Novel, Efficient Maximum-Likelihood Algorithm." *Genetics* 200, no. 1: 285–293.
- Jackson, J. A., L. Laikre, C. S. Baker, K. C. Kendall, and Genetic Monitoring Working Group. 2012. "Guidelines for Collecting and Maintaining Archives for Genetic Monitoring." *Conservation Genetics Resources* 4: 527–536.
- Jorde, P. E., and N. Ryman. 1995. "Temporal Allele Frequency Change and Estimation of Effective Size in Populations With Overlapping Generations." *Genetics* 139, no. 2: 1077–1090.
- Jorde, P. E., and N. Ryman. 2007. "Unbiased Estimator for Genetic Drift and Effective Population Size." *Genetics* 177, no. 2: 927–935.
- Kamath, P. L., M. A. Haroldson, G. Luikart, D. Paetkau, C. Whitman, and F. T. Van Manen. 2015. "Multiple Estimates of Effective Population Size for Monitoring a Long-Lived Vertebrate: An Application to Y Ellowstone Grizzly Bears." *Molecular Ecology* 24: 5507–5521.
- Krimbas, C. B., and S. Tsakas. 1971. "The Genetics of *Dacus oleae*. V. Changes of Esterase Polymorphism in a Natural Population Following Insecticide Control-Selection or Drift?" *Evolution* 25: 454–460.
- Lewontin, R. C., and J. Krakauer. 1973. "Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms." *Genetics* 74, no. 1: 175–195.
- Li, X., C. Zhu, Z. Lin, et al. 2011. "Chromosome Size in Diploid Eukaryotic Species Centers on the Average Length With a Conserved Boundary." *Molecular Biology and Evolution* 28, no. 6: 1901–1911.
- Luikart, G., J. M. Cornuet, and F. W. Allendorf. 1999. "Temporal Changes in Allele Frequencies Provide Estimates of Population Bottleneck Size." *Conservation Biology* 13, no. 3: 523–530.
- Luikart, G., N. Ryman, D. A. Tallmon, M. K. Schwartz, and F. W. Allendorf. 2010. "Estimation of Census and Effective Population Sizes: The Increasing Usefulness of DNA-Based Approaches." *Conservation Genetics* 11: 355–373.
- Moyle, P. B., B. Herbold, D. E. Stevens, and L. W. Miller. 1992. "Life History and Status of Delta Smelt in the Sacramento-San Joaquin Estuary, California." *Transactions of the American Fisheries Society* 121, no. 1: 67–77.
- Mwima, R., T. Y. J. Hui, J. K. Kayondo, and A. Burt. 2024. "The Population Genetics of Partial Diapause, With Applications to the Aestivating Malaria Mosquito *Anopheles Coluzzii*." *Molecular Ecology Resources* 24, no. 4: e13949.
- Nei, M., and F. Tajima. 1981. "Genetic Drift and Estimation of Effective Population Size." *Genetics* 98, no. 3: 625–640.
- Palstra, F. P., and D. J. Fraser. 2012. "Effective/Census Population Size Ratio Estimation: A Compendium and Appraisal." *Ecology and Evolution* 2, no. 9: 2357–2365.
- Pollak, E. 1983. "A New Method for Estimating the Effective Population Size From Allele Frequency Changes." *Genetics* 104, no. 3: 531–548.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ruzzante, D. E., G. R. McCracken, S. Parmelee, et al. 2016. "Effective Number of Breeders, Effective Population Size and Their Relationship With Census Size in an Iteroparous Species, *Salvelinus fontinalis*." *Proceedings of the Royal Society B: Biological Sciences* 283, no. 1823: 20152601.
- Schwartz, M. K., G. Luikart, and R. S. Waples. 2007. "Genetic Monitoring: A Promising Tool for Conservation and Management." *Trends in Ecology & Evolution* 22: 25–33.
- Tempel, T. L., T. D. Malinich, J. M. Burns, A. Barros, C. E. Burdi, and J. A. Hobbs. 2021. "The Value of Long-Term Monitoring of the San Francisco Estuary for Delta Smelt and Longfin Smelt." *California Fish & Game* 107: 148–171.
- Therkildsen, N. O., J. Hemmer-Hansen, T. D. Als, et al. 2013. "Microevolution in Time and Space: SNP Analysis of Historical DNA Reveals Dynamic Signatures of Selection in Atlantic Cod." *Molecular Ecology* 22, no. 9: 2424–2440.
- Thurfjell, H., L. Laikre, R. Eklom, S. Hoban, and P. Sjögren-Gulve. 2022. "Practical Application of Indicators for Genetic Diversity in CBD Post-2020 Global Biodiversity Framework Implementation." *Ecological Indicators* 142: 109167.
- United States Fish and Wildlife Service. 2020. *Delta Smelt Supplementation Strategy*, 55. CA: Sacramento.
- United States Fish and Wildlife Service. 2022. *BY2021 ERTT Summary of Activities*, 55. California: Sacramento.
- Van Rossum, F., and O. J. Hardy. 2022. "Guidelines for Genetic Monitoring of Translocated Plant Populations." *Conservation Biology* 36, no. 1: e13670.
- Wang, J. 2001. "A Pseudo-Likelihood Method for Estimating Effective Population Size From Temporally Spaced Samples." *Genetics Research* 78, no. 3: 243–257.
- Wang, J. 2009. "A New Method for Estimating Effective Population Sizes From a Single Sample of Multilocus Genotypes." *Molecular Ecology* 18, no. 10: 2148–2164.
- Waples, R. S. 1989. "A Generalized Approach for Estimating Effective Population Size From Temporal Changes in Allele Frequency." *Genetics* 121: 379–391.
- Waples, R. S. 2005. "Genetic Estimates of Contemporary Effective Population Size: To What Time Periods Do the Estimates Apply?" *Molecular Ecology* 14: 3335–3352.
- Waples, R. S. 2006. "A Bias Correction for Estimates of Effective Population Size Based on Linkage Disequilibrium at Unlinked Gene Loci." *Conservation Genetics* 7: 167–184.
- Waples, R. S. 2022. "What Is N_e , Anyway?" *Journal of Heredity* 113: 371–379.
- Waples, R. S., and C. Do. 2008. "*LdNe*: A Program for Estimating Effective Population Size From Data on Linkage Disequilibrium." *Molecular Ecology Resources* 8: 753–756.
- Waples, R. S., M. Masuda, and J. Pella. 2007. "*SALMONNb*: A Program for Computing Cohort-Specific Effective Population Sizes (N_e) in Pacific Salmon and Other Semelparous Species Using the Temporal Method." *Molecular Ecology Notes* 7: 21–24.
- Waples, R. S., M. M. Masuda, M. E. F. LaCava, and A. J. Finger. 2024. "*MaxTemp*: Software to Maximize Precision of the Temporal Method for Estimating N_e in Genetic Monitoring Programs." <https://doi.org/10.5281/zenodo.14213748>.

Waples, R. S., R. K. Waples, and E. J. Ward. 2022. "Pseudoreplication in Genomics-Scale Datasets." *Molecular Ecology Resources* 22: 503–518.

Waples, R. S., and M. Yokota. 2007. "Temporal Estimates of Effective Population Size in Species With Overlapping Generations." *Genetics* 175: 219–233.

Whiteley, A. R., J. A. Coombs, M. Cembrola, et al. 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, no. 14: 3585–3602.

Williamson, E. G., and M. Slatkin. 1999. "Using Maximum Likelihood to Estimate Population Size From Temporal Changes in Allele Frequencies." *Genetics* 152: 755–761.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.