

# Estimating effective population size of disease vectors: a critical assessment of applications and performance

Tiago Antao\*, Ian M. Hastings, Gordon Luikart, Martin J. Donnelly

## Abstract

Estimation of the contemporary effective population size ( $N_e$ ) is increasingly conducted in insect disease vectors.  $N_e$  estimation is used to assess likely changes in genetic diversity and also to evaluate the impact of control measures like insecticide-treated bed nets (ITNs) or indoor residual spraying (IRS) in reducing vector populations. We evaluated the performance of the most commonly used  $N_e$  estimator based on F-statistics which uses two temporally spaced samples and compared it to two recent estimators based on likelihood (another 2-sample temporal method) and linkage disequilibrium (a 1-sample method). We simulated three different demographies based on realistic parameters for common vector species: a constant population size model; bottlenecks to simulate control measures and a novel model with sinusoidal demography to simulate vector populations which typically expand seasonally with dry and wet seasons. Results show that the sampling strategies used in most empirical studies ( $\sim 60$  individuals and  $\sim 10$  microsatellite loci per temporal sample) are not sufficient to estimate  $N_e$  with precision. We suggest that an increase in temporal spacing between samples or an increase to  $\sim 50$  loci might be sufficient to provide sufficient precision. The likelihood method outperforms the classic  $F_k$  moments estimator in most cases. In fluctuating population scenarios the temporal and LD methods provide qualitatively different estimates. For example the LD method is very sensitive to immediate demographic changes whereas the temporal

---

\*Corresponding author. Address: Liverpool School of Tropical Medicine, Liverpool, UK. E-mail: Tiago.Antao@liverpool.ac.uk

methods provide an average over several generations. This suggests that each class of methods may have different applications as the LD method can detect seasonal patterns and control interventions whereas the temporal estimators quantify trends over a longer period of time.

**Keywords:** effective population size, statistical power, microsatellites, insect disease vectors, population control, seasonality

## 1 Introduction

The effective population size ( $N_e$ ) provides a measure of the rate of random genetic change in populations caused by genetic drift [Charlesworth, 2009], and the relative efficiency of natural selection in the face of drift.  $N_e$  is also a fundamental factor determining population viability as larger  $N_e$  entails greater population genetic variability which is paramount for species survival and adaptation. Estimation of  $N_e$  is increasingly conducted for insect disease vectors such as the *Anopheles* vectors of malaria and the *Glossina* vectors of African trypanosomiasis. Commonly the purpose of these studies is to determine whether insecticide-based control measures have successfully reduced the contemporary  $N_e$  [Wondji et al., 2005] or to investigate whether populations undergo seasonal contractions [Simard et al., 2000].

Indeed, one of the most widely used estimators of  $N_e$  across all taxa was developed to study the impact of insecticide resistance [Krimbas and Tsakas, 1971]. This moments based method of contemporary  $N_e$  estimation was further developed by [Nei and Tajima, 1981, Pollak, 1983, Waples, 1989] and is based on obtaining at least two samples displaced over time (generations) and estimating the temporal variance in allele frequencies between them.

Luikart et al. [1999] demonstrated that this temporal method was far more powerful for detecting population declines than tests based on loss of alleles or heterozygosity for detecting population declines. Effective population size should not be confused with census size ( $N_c$ ), i.e. the total number of individuals in a population at any given time.  $N_c$  is normally larger than  $N_e$  for wildlife species with  $\frac{N_e}{N_c}$  ratios estimated around 0.10

[Frankham, 1995, Kalinowski and Waples, 2002], although there is only one known study for this ratio in parasite vectors [Solano et al., 2009] where ratios calculated exhibit a large variance.

The second most widely used class of estimators of contemporary  $N_e$  is the single-sample estimator based the linkage (gametic) disequilibrium (LD) method [Hill, 1981]. Waples and Do [2010] showed that LDNe can provide precise estimates of  $N_e$  in constant-sized populations with non-overlapping generations by using 10–20 microsatellite loci (5–10 alleles/locus) and samples of at least 25–50 individuals, if the effective population size is less than approximately 500.

While the temporal estimator has been studied both with constant sized populations and bottlenecks [see e.g. Luikart et al., 1999, Antao et al., 2011] for low  $N_e$ , there has never been a critical assessment of its usage alone or in comparison to single sample methods (LDNe) in the context of vector biology. This assessment is increasingly needed given the importance of vector-based infectious diseases, increasing risks of emerging disease following environmental changes, seasonality and the need to evaluate of control interventions to reduce vector population size.

## Standard assumptions for insect vector studies

The assumptions commonly made for vector population studies are that individuals are sampled without replacement prior to reproduction (Dyer et al. [2009] and others) this is plan II sampling of [Waples, 1989]. Following work of [Lehmann et al., 1998] most of authors have take a conservative estimate of 12 generations per year for the African malaria vector *Anopheles gambiae* which allows estimates to be compared between studies. Researchers on South East Asian anophelines have used a value of 10 generations per year [Walton et al., 2000]. Consequently the reported values may underestimate the real  $N_e$  if there are more generations per year. Authors also generally assume that allele frequency change is solely attributable to genetic drift [Dyer et al., 2009], whereas as selection or sampling bias due to substructure could also lead to allele frequency changes. The final assumption is of constant population size whereas many insect vectors exhibit

extreme seasonal variation in census size with populations increasing following the onset of favourable conditions. For tropical mosquito species this is often the onset of rainy season [Charlwood et al., 1995, TAYLOR et al., 1993].

Studies of  $N_e$  also use different sampling strategies with varying number of individuals sampled, number of loci, and especially the temporal spacing between samples. It is also assumed that there is independent sampling of individuals and independent (unlinked) loci Dyer et al. [2009]. The former assumption may be difficult to defend giving the increasing evidence of stratification in vector populations [Weetman et al., 2010]. A representative sample of  $N_e$  studies in vectors is shown on Table 1. We will concentrate on studies reporting high  $N_e$  (i.e. above 100), because the behaviour of estimates with low  $N_e$  (like the values reported in [Solano et al., 2009]) has been widely studied [e.g Tallmon et al., 2010, Waples and Do, 2010, Berthier et al., 2002].

Alternative estimators of contemporary  $N_e$  have been proposed, either (i) updated versions of the temporal method using Maximum Likelihood (ML), such as the ones implemented in the MLNE [Wang, 2001, Wang and Whitlock, 2003] or TM3[Berthier et al., 2002] applications or (ii) completely different approaches based on a single sample and linkage disequilibrium (LD) [Hill, 1981, Waples, 2006] as implemented in LDNe [Waples and Do, 2008]. MLNE and TM3 were shown to have better performance than the standard temporal method for constant sized populations [Wang, 2001, Berthier et al., 2002] but doubts were raised if that was the case with MLNE in bottleneck scenarios [Antao et al., 2011]. The LD method has been compared to the temporal method for constant population size scenarios [Waples and Do, 2010] and bottleneck detection [England et al., 2010, Antao et al., 2011].

Here we present a computational study, using individual-based forward-time population genetic simulations, evaluating several contemporary  $N_e$  estimators using realistic demographies for insect vectors. The rationale for this study is two-fold:

1. We critically appraise existing published studies of  $N_e$  in insect vectors and determine whether
  - The number of individuals and loci are sufficient to provide an unbiased esti-

mate of  $N_e$  with a reasonable confidence interval.

- The temporal spacing between samples spans enough generations to provide a accurate estimation of  $N_e$ .
- The temporal method will allow us to estimate the impact of vector control or seasonality on vector population size.

2. We also provide guidelines for future studies of  $N_e$  by

- Investigating if recent approaches to estimate  $N_e$  perform better than the original moments based temporal method.
- Studying the impact of realistic vector demographics on the estimators.
- Suggesting sampling strategies, capable of providing sufficient precision when studying the impact of control measures and seasonality.
- Understanding which estimator is most appropriate for studying seasonality and/or control impacts.

## 2 Methods

We start by presenting an overview of the standard temporal method, followed by an introduction to maximum likelihood temporal estimator and the LD method. We then describe the simulations and sampling strategies.

### 2.1 Moment-based $F_k$ temporal method

For the temporal method we implemented the  $N_e$  estimator from Waples [1989] based on Nei and Tajima [1981] and Krimbas and Tsakas [1971]:

$$\hat{N}_e = \frac{t}{2 \left[ \hat{F}_k - \frac{1}{2S_0} - \frac{1}{2S_t} \right]} \quad (1)$$

Where  $t$  is the time between generations,  $S_0$  is the sample size (number of individuals) at the reference, pre-bottleneck point and  $S_t$  at the post-bottleneck generation being

considered. This is the estimator for plan II of Waples [1989] (sampling destructively before reproduction). Though in our simulations we sample non-destructively (plan I), the difference between estimators with high  $N_c$  (like the values simulated here) is expected to be low [Waples, 1989]. The plan II estimator has been extensively evaluated and its usage will allow for comparative analysis, furthermore it is the estimator commonly used for vectors.

The  $F_k$  estimator is implemented for each locus ( $l$ ) as [Pollak, 1983]:

$$\hat{F}_k^l = \frac{2}{K-1} \sum_{i=1}^K \frac{(f_{ri} - f_{ti})^2}{f_{ri} + f_{ti}} \quad (2)$$

Where  $K$  is the number of alleles at the current loci,  $f_{ri}$  is the frequency of allele  $i$  at the reference time and  $f_{ti}$  is the frequency of allele  $i$  at the current time. The  $F_k$  value used in the  $N_e$  estimator will be the weighted arithmetic mean of all locus  $F_k$  estimators (equation 2), the weight being the number of alleles.

Confidence Intervals (CI) on  $\hat{F}$ , which can be used to calculate the CI of  $\hat{N}_E$ , were computed as follows [Waples, 1989, Sokal and Rohlf, 1995, Luikart et al., 1999]:

$$\alpha(1-\alpha)CI \text{ for } \hat{F}_k^l = \left[ \frac{n\hat{F}_k^l}{\chi_{\alpha/2[n]}^2}, \frac{n\hat{F}_k^l}{1 - \chi_{1-\alpha/2[n]}^2} \right] \quad (3)$$

Where  $1 - \alpha$  is the proportion of CIs containing the real  $N_e$  and  $n$  is the number of independent alleles given by:

$$n = \sum_{i=1}^l (K_i - 1) \quad (4)$$

Where  $K_i$  is the number of alleles of locus  $i$ .

The coefficient of variation (CV) of the temporal estimator was presented in Pollak [1983]:

$$CV(\hat{N}_e) \approx \sqrt{\frac{2}{n}} \left[ 1 + \frac{2N_e}{tS} \right] \quad (5)$$

Where  $S$  is the number of individuals sampled. The CV, a measure of dispersion,

suggests that precision is increased if either the spacing between samples or the number of alleles increases. The estimator is also expected to lose precision with large real  $N_e$ , as the actual value is in the numerator of the CV.

This moment-based estimator is known to be biased upwards [Waples, 1989, Berthier et al., 2002] and that rare alleles are largely responsible for overestimation so we also pooled (binned) all alleles with frequency below 2% into a single class. Results from binning were compared with the standard (without binning) estimator.

## 2.2 Likelihood-based temporal estimator

Several estimators have been proposed which use temporal sampling and maximum-likelihood (ML). For example, the ML method in Berthier et al. [2002] has been used in Dyer et al. [2009] to estimate  $N_e$  in a *Glossina palpalis palpalis* population in Equatorial Guinea. ML based estimators should provide better precision than moment-based estimators because they use more information from the data [Edwards, 1972] and the ML method used in Dyer et al. [2009] has been shown to perform better with very low  $N_e$  (i.e. 20) and when dealing with rare alleles [Berthier et al., 2002]. This multiallelic method is based on coalescent simulation. Like most ML methods, extensive testing is computationally costly in terms of time. Strictly speaking this method is Bayesian as a maximum  $N_e$  prior has to be supplied. This estimator is implemented in the TM3 application. The parameters for each estimate were extracted from 20,000 coalescent simulations where the  $N_e$  estimator was capped at 15,000. The following summary statistics are computed: mode and the 0.025 and 0.975 quantiles (giving a 95% credible interval). There is no known CV for this estimator, but the same variables that influence the moments-based estimator (real  $N_e$ , time between samples, number of alleles and sample size) are expected to impact this estimator in qualitatively similar ways.

## 2.3 Linkage disequilibrium estimator

Linkage disequilibrium can be used to estimate effective population size as its variance is a function of  $N_e$  and sample size. [Hill, 1981] noted that the variance of LD estimates

across loci is a function of both the effective population size and sampling strategy and proposed an estimator based on LD. This estimator has been shown to be upwardly biased if the sample size is smaller than the true  $N_e$  [England et al., 2006] and a bias correction has been proposed [Waples, 2006]. The LD method has one main clear advantage over temporal approaches: it requires only a single sample. The LD method implemented in LDNe [Waples and Do, 2008] has been compared to the moment based temporal method for equilibrium (i.e., constant population size) scenarios [Waples and Do, 2010] and bottlenecks [Antao et al., 2011]. Evaluations of performance are also given in Tallmon et al. [2010] and England et al. [2010]. The CV for this estimator is [Hill, 1981, Waples and Do, 2010]:

$$CV_{LD}(\hat{N}_e) \approx \sqrt{\frac{2}{n'}} \left[ 1 + \frac{3N_e}{S} \right] \quad (6)$$

Where  $n'$  is:

$$n' = \sum_{i=1}^{L-1} \sum_{j=i+1}^L (K_i - 1)(K_j - 1) \quad (7)$$

Both the coefficient of variation and computational studies suggest that, like the temporal method, the LD estimator has increased absolute precision for low real  $N_e$  and larger sample sizes ( $n'$  increases with both number of loci and alleles per loci).

Point estimates and 95% confidence intervals (parametric) are computed using only alleles with a frequency of 2% or more in order to correct for upward bias. This correction is reported to provide an acceptable balance between precision and bias [Waples and Do, 2010] for the sample strategies tested (when  $S > 25$ ).

## 2.4 Simulations and demographies

We conducted simulations using the forward-time, individual based simulator simuPOP [Peng and Kimmel, 2005]. Simulations were performed using a Wright-Fisher model with separate sexes, random mating (average sex ratio of 1) and discrete, non-overlapping generations. This makes  $\frac{N_c}{N_e} \approx 1$ . Each demographic scenario was replicated 1,000 times.



Simulations had a burn-in phase of at least 10 generations in order to approximate mean observed heterozygosity with realistic values (below 0.8). Longer burn-in periods were also tested, but results were qualitatively unchanged. The genome simulated includes 100 neutral, independent microsatellite loci initialized with a Dirichelet distribution (10 initial alleles per locus exhibiting a mean of 8 after burn-in, approximating the conditions on table 1) and a mutation rate of  $10^{-4}$  using a stepwise mutation model [Lehmann et al., 1998]. Simulations with a larger number of starting alleles (up to 20) were also conducted.

Three different demographies were tested: i) a standard demography of constant size, ii) a bottleneck potentially imposed by a transmission control measure such as ITNs or IRS and iii) a novel model where the population size varies with a cosine function in order to model vector seasonality (figure 4). The constant scenario was run with a  $N_c$  of 200, 400, 800, 1000, 2,000 and 4,000. The bottleneck scenario started with a  $N_c$  of 5,000 or 2,000 which were then reduced ten-fold. The seasonal scenario was based on  $A\cos(\frac{2t\pi}{12}) + B$  where  $t$  is the generation, and  $A$  and  $B$  are parametrized according to the demography. The function above implicitly defines a period of 12 generations, based on 12 generations per year for *A. gambiae*. For  $A$  and  $B$  we used the parameters  $A = 500$ ,  $B = 700$ , making the minimum  $N_c$  of 200 and the maximum of 1,200.

To perform the  $N_e$  estimation we sampled 60 individuals per generation and 10 microsatellite loci, in line with existing studies. We also studied the impact of doubling the sample size (120 individuals) and tested different numbers of loci sampled (20, 50 and 100). For the both temporal methods we tested different temporal distances between sampling ranging from 4 generations to 100 (this does not apply to the LD method as it is based on a single sample). Due to the extreme computational cost of the ML method we only studied 100 replicates (instead of 1,000) per scenario.

## 3 Results

### 3.1 Moment based temporal method

#### Bias

For a time span of 4 the real  $N_e$  is always below the lower quartile of the distribution of 1,000 point estimates, and the upper quartile is always above 3 times the  $N_e$  value. For  $N_e$  of 1,000 and 2,000 the lower quartile occasionally includes the real value, but only due to increased lack of precision. For time spans of 12 and 24 generations the median point estimate (among 1000 replicates) is always above the real value (i.e. there is upward bias), but in most cases the upper quartile is below 3 times the true value. As expected, sampling more individuals and/or more loci provides more precise estimates. Sampling more individuals appears to be slightly more informative than sampling more loci as the precision of sampling 60 individuals and 20 loci is slightly lower than the precision with 120 individuals and 10 loci. Figure 1 shows the distribution of point estimates for spans of 4, 12 and 24 generations using different sampling strategies. Binning, i.e. pooling all alleles with a frequency below 2% in a single class in order to reduce upward bias is useful for loci with many alleles but rarely reduces bias substantially.

### **Precision**

For an  $N_e$  of 2,000 the confidence intervals of point estimates is large, the harmonic mean of the upper CI is always above 6,000 ( $3N_e$ ) assuming that the time span between sampled generations is below 25 (figure 2). The CI for an  $N_e$  of 1,000, the common sampling policy with 60 individuals and 10 loci will also be above 3 times  $N_e$  for all time spans. For this sampling strategy a time span of one year with *A. gambiae* is not enough to have a upper confidence interval below 3 times the  $N_e$  values simulated (including a  $N_e$  of 500). The upper confidence interval will also never be below 3 times the  $N_e$  for all sampling strategies if the time span between generations is below 8 irrespective of  $N_e$ . Again, sampling more individuals provides slightly better results than the proportional sampling of more loci. Figure 2 plots the harmonic mean for the point estimate and 95% confidence intervals with time spans between samplings up to 24 generations using different sampling strategies.

### 3.2 Comparison of methods

In order to compare the 3 methods used to estimate contemporary  $N_e$ , we show on figure 3 a box plot for point estimates and also the harmonic mean of 95% confidence intervals (or the credible intervals for the ML method). For the temporal methods we include 2 time spans (4 and 24 generations). The ML method is more precise than the original moments based estimator assuming equal time spans with the exception of the typical sampling strategy and a time span of 4. In terms of bias the LD method is always less biased than any of the temporal methods for an  $N_e$  of 500. Temporal methods are in most cases less precise than LDNe for a time span of 4 and always more precise with time spans of 24, reinforcing the importance of the time between samples. A time span of 4 will always produce imprecise results unless the ML method is used with 60 individuals and 50 loci. While TM is more precise, it is also severely biased downwards with an  $N_e$  of 1,000 as the upper quartile is below 1,000.

### 3.3 Fluctuation in population size

For seasonal models the temporal estimators and the LD method show different behaviour. The LD method will mostly estimate the  $N_e$  value of the previous generation whereas the temporal methods will estimate a mostly stable average between maximum and minimum  $N_e$ . This makes the LD estimator vary from generation to generation while the temporal estimators are much more stable (figure 4 top). While the median estimation of the LD method will approach the  $N_e$  of the previous generation, the precision of the estimations will produce overlaps (figure 4 bottom), the precision will be especially low when the estimator is based on higher  $N_e$ s. Only at the extremes ( $N_e$  of 200 and 1,200) will the overlap be minimal.

The LD method is able to detect a bottleneck as soon as one generation after it occurs for an  $N_e$  of 200 (figure 5) and 500. The temporal methods proceed much slowly (i.e. are much less sensitive) as for nine generations after the bottleneck, the moments based estimator is still above 3 times  $N_e$  for both bottleneck scenarios. The ML estimator has much lower upward bias (i.e. is more influenced by the sample with lower  $N_e$ ) than the

moments based version. Again like in the seasonal model, the temporal methods average the trend spanning the sampling period while the LD method reflects the state at the sampling time.

## 4 Discussion

With the temporal methods, very small timespans between samples severely decreases precision and enlarges CIs. A clear example of this effect can be seen in [Simard et al., 2000] where all estimates made with 4 months of separation (assumed conservatively to be approximately 4 generations for *A. gambiae*) include infinity in the confidence interval as expected. In the same study, estimations with more than 3 years of interval (approximately 40 generations) do provide much tighter confidence intervals (never including infinity). The ML method, while more precise with a timespan of 4 either shows extreme upward bias and no precision with small sampling strategies (rendering it useless) or strong downward bias, with tight precision with larger samples (60 samples and 50 loci). With a timespan of 24 the ML method is both more precise and less biased than the moments based method. Our results suggest that the temporal methods are not useful for any purpose if the time span is low.

The classic moments based method produces estimates that are normally biased upwards. The strategy of binning rare alleles is not enough to eliminate the bias, though it improves results if loci with many alleles are included (results not shown). Both the LD and ML methods are less biased than the moments based method. It is not clear that binning rare alleles with the ML method (as done in Dyer et al. [2009]) is a good strategy as this method is known to perform better with rare alleles [Berthier et al., 2002] and our results suggest that the estimator is normally biased downwards, thus any binning might compound the problem. The ML method is only strongly biased upwards when the sample size is too low (e.g with 35 individuals and 12 loci as in Dyer et al. [2009]). Indeed, the results in Dyer et al. [2009] where the ML estimation is higher than the classic estimate suggests that the sampling size might have been insufficient. This relationship between estimators happens with small sample size (figure 3) and the sample size per

time point was indeed very low. Sample sizes of 30 were also shown to be insufficient for the LD method [Tallmon et al., 2010], even with much lower real  $N_e$ .

Thus likelihood approach also provides more precision and tighter confidence intervals than the classic estimator in almost all cases. Only with small time spans (4 generations) and few loci (10 or 20) the precision of the ML estimator is worse but this “advantage” is mostly theoretical as the precision of the classic estimator, while better with small time spans is still very bad, i.e., while the moments based estimator performs, in theory, better, it is still unusable.

For very large time spans both temporal estimators provide more precise estimates than the single-sample LD method. This means that the span between samples is a crucial parameter in deciding which method to use. From a practical perspective the temporal methods require twice the sampling effort as two time points are needed to make an estimate. In deciding which method to use researchers should consider not only the possible time span between samples in the temporal method, but also the economic issue of using one or two time samples: much more precision can be gained from, having a single sample with 120 and applying the single-sample LD method instead of having two time samples of 60 individuals each.

Higher real  $N_e$  decreases precision with all estimators, therefore the empirical temporal estimates sampled a year apart (circa 12 generations) in the vector control study of Pinto et al. [2003] either include infinity in the upper confidence interval (with a high  $\hat{N}_e$  of 1,078) in one case or even the point estimate is infinite in another case. The 24 month estimate of 1,457 has a upper confidence interval of almost one order of magnitude above (13,677). Such estimates, including infinite in the confidence intervals (and even in point estimates) make any inference of the impact of control measures unreliable at best.

For most scenarios tested precision is more problematic than bias. Though our results show bias that can go up to 50% with the temporal method, precision can vary above one order of magnitude between the lower and upper quartile. Furthermore, our simulation assumption of Wright-Fisher equilibrium (i.e.  $N_c = N_e$ ) is expected to have maximum bias, whereas for more realistic relationships between  $N_c$  and  $N_e$ , bias can be negligible

[Waples, 1989]. For realistic time spans (i.e. above 4), the sampling strategy has more influence on precision than on bias, therefore it is possible to vastly improve precision with a better sampling policy.

Our results suggest that the common sampling policy of 60 individuals and 10 loci is very unprecise in almost all cases and for any interesting interpretation of the estimates. Only with long time spans between temporal samples and where  $N_e$  is less than 500 are the estimates precise. In theory increasing the number of individuals will achieve a slightly higher increase in precision than increasing the number of loci. However, given the problems with collecting some of these vector species and the advent of new high throughput genotyping platforms, a more practical suggestion in most cases may be to increase the number of loci. Using 60 individuals and 50 loci will allow for precise estimates, if  $N_e$  is less than 1,000, though if the expected value of  $N_e$  is close to 1,000 increased spacing between samples (above 24 generations) may be required. Next generation sequencing will allow the usage of thousands of SNP markers and further research should consider the performance of estimators in such sampling scenarios with thousands of bi-allelic loci with varying linkage disequilibrium. Our results also suggest that, for both methods sampling more individuals yields a bigger increase in precision than sampling more loci. This is consistent with results for the ML temporal method [Berthier et al., 2002]. For the LD method most other studies [England et al., 2010, Antao et al., 2011] (but not all, see Tallmon et al. [2010]) suggest that indeed more individuals provide more information.

When the population is not constant the temporal and LD methods have qualitatively different behaviours. The LD method is extremely sensitive to the  $N_e$  of the previous 1 or 2 generations whereas the temporal methods “smooth” the ongoing demographic processes (figures 4 and 5). This is not a suggestion that one class of method is “better” or “worse” than the other, only that they have different applications: the LD method is better suited for early detection of bottlenecks (e.g. population reducing interventions) or to study seasonality whereas the temporal estimators provide a better picture of the “average” population size. It should be noted however that a less naive interpretation of LD results suggests that it might not be applicable to study seasonality as the confidence intervals

of the seasonal point estimates overlap (as with higher values of  $N_e$  the precision drops). Even if the point estimate distribution could be tightened (by increasing sample size, especially with more individuals sampled as the LD method is substantially more sensitive to individuals than loci [Antao et al., 2011]), the confidence intervals might still overlap for a fine-grained (monthly) estimation, though detection of extremes will probably be feasible. Any previous conclusion made about seasonality and control measures using temporal estimators is thus fraught with uncertainty. While we used the temporal ML method in Berthier et al. [2002] as it was used in a vector biology study with large  $N_e$  [Dyer et al., 2009], more recent ML temporal based methods developed to detect bottlenecks [Beaumont, 2003] should be investigated in the context of vector biology to assess their performance to detect control interventions and perhaps seasonality.

Simard et al. [2000] suggests [based on Nei and Tajima, 1981, Pollak, 1983] that, as the moments-based estimator approximates the harmonic mean of the effective population sizes, it is dominated by the smallest value a fact also highlighted by O’Ryan et al. [1998]. Our findings are not inconsistent with this statement but we note the following: (i) While this effect is visible, the temporal estimators are still distant from the contemporary  $N_e$ , (ii) if there is an expansion, the estimator is not useful and (iii) as the time distance between the bottleneck and the second sample increases the estimator will tend to approach the contemporary value. This latter effect is possibly caused by drift that is not compensated by the introduction of new mutations.

Testing the ML estimator in a wide array of scenarios is not feasible due to the extremely high computational cost of the TM3 application (common to most likelihood approaches), where even the manuscript describing the method [Berthier et al., 2002] suggests that the application slow speed precluded the evaluation of many scenarios. While running a single instance of the application is computationally cheap, running thousands of evaluations is prohibitive. At the risk of sparking controversy we raise the following question: if extensive testing of likelihood applications is computationally unfeasible and only limited *ad-hoc* tests are possible, can we often trust the results of ML approaches?

Some simulation assumptions might require further investigation in the particular

context of vector biology: while random mating might be a reasonable assumption with Anophelines, it is less clear that it is acceptable for other vector species such as *Glossina* as polyandry and sperm competition is widespread with insects [Simmons, 2001, Tripet et al., 2001]. Age structure might be of particular importance with regards to seasonality estimates with the LD method as the signal from previous generations might “smooth” the estimation curve. On the other hand, as the ratio between  $N_e$  and  $N_c$  is likely well below 1, bias with the temporal method will most probably be much lower than reported here. If a sample is mostly originating in a small breeding site where close genetic relatedness between individuals exists, the results above might also not be applicable though in this case the problem resides more in the sample itself as it might not be representative of the whole population.

## 5 Conclusion

We suggest that most existing empirical studies might require re-interpretation of results. It is not clear that common sampling strategies are sufficient to reliably estimate  $N_e$  in general or to detect the impact of control measures and to estimate seasonality of vectors. The temporal and LD methods have different applications, while the temporal methods provide an average measure of  $N_e$  over the sampled period, the LD method is extremely sensitive (barring age-structure effects) to sudden changes in population size. The LD method can be used to assess the success of control interventions (bottlenecks) and may help infer seasonality patterns of fluctuation in effective population size.

While previous studies often lack sufficient sample sizes (and in some cases enough time between sampled generations), contemporary  $N_e$  estimators can reliably be used to infer population size and evaluate control interventions as long as samples (of loci and individuals) and spacing between generations are sufficiently large for the real  $N_e$  and the correct estimators are applied, which will be dependent on the question being asked, and to carefully interpret results.



## 6 Acknowledgments

TA was supported by research grant SFRH/BD/30834/2006 from Fundacao para a Ciencia e Tecnologia, Portugal.

## References

- Tiago Antao, Andrs Prez-Figueroa, and Gordon Luikart. Early detection of population declines: high power of genetic monitoring using effective population size estimators. *Evolutionary Applications*, 4:144–154, 2011. ISSN 1752-4571. doi: 10.1111/j.1752-4571.2010.00150.x. URL <http://dx.doi.org/10.1111/j.1752-4571.2010.00150.x>.
- M.A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139, 2003.
- P. Berthier, M.A. Beaumont, J.M. Cornuet, and G. Luikart. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, 160(2):741, 2002.
- Brian Charlesworth. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3):195–205, Mar 2009. doi: 10.1038/nrg2526.
- JD Charlwood, J. Kihonda, S. Sama, PF Billingsley, H. Hadji, JP Verhave, E. Lyimo, PC Luttikhuisen, and T. Smith. The rise and fall of *Anopheles arabiensis*(Diptera: Culicidae) in a Tanzanian village. *Bulletin of entomological research*, 85(01):37–44, 1995. ISSN 1475-2670.
- NA Dyer, A. Furtado, J. Cano, F. Ferreira, M. Odete Afonso, N. NDONG-MABALE, P. NDONG-ASUMU, S. CENTENO-LIMA, A. Benito, D. Weetman, et al. Evidence for a discrete evolutionary lineage within Equatorial Guinea suggests that the tsetse fly *Glossina palpalis palpalis* exists as a species complex. *Molecular Ecology*, 18(15): 3268–3282, 2009. ISSN 1365-294X.

- A.W.F. Edwards. *Likelihood: an account of the statistical concept of likelihood and its application to scientific inference*, volume 235. Cambridge University Press, 1972.
- Phillip England, Gordon Luikart, and Robin Waples. Early detection of population fragmentation using linkage disequilibrium estimation of effective population size. *Conservation Genetics*, 11:2425–2430, 2010. ISSN 1566-0621. URL <http://dx.doi.org/10.1007/s10592-010-0112-x>. 10.1007/s10592-010-0112-x.
- P.R. England, J.M. Cornuet, P. Berthier, D.A. Tallmon, and G. Luikart. Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conservation Genetics*, 7(2):303–308, 2006. ISSN 1566-0621.
- R. Frankham. Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, 66(02):95–107, 1995. ISSN 1469-5073.
- William G. Hill. Estimation of effective population size from data on linkage disequilibrium. *Genetics Research*, 38(03):209–216, 1981.
- S.T. Kalinowski and R.S. Waples. Relationship of effective to census size in fluctuating populations. *Conservation Biology*, 16(1):129–136, 2002. ISSN 1523-1739.
- Costas B. Krimbas and Spyros Tsakas. The genetics of *dacus oleae*. v. changes of esterase polymorphism in a natural population following insecticide control-selection or drift? *Evolution*, 25(3):454–460, 1971.
- T. Lehmann, W.A. Hawley, H. Grebert, and F.H. Collins. The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Molecular Biology and Evolution*, 15(3):264, 1998. ISSN 0737-4038.
- Gordon Luikart, Jean-Marie Cornuet, and Fred W. Allendorf. Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology*, 13(3):523–530, 1999. ISSN 08888892. URL <http://www.jstor.org/stable/2641867>.
- Masatoshi Nei and Fumio Tajima. Genetic drift and estimation of effective population size. *Genetics*, 98(3):625–640, 1981.

- C. O’Ryan, E.H. Harley, M.W. Bruford, M. Beaumont, R.K. Wayne, and M.I. Cherry. Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Animal Conservation*, 1(02):85–94, 1998. ISSN 1469-1795.
- B. Peng and M. Kimmel. simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, 21(18):3686–3687, 2005. doi: 10.1093/bioinformatics/bti584.
- J. Pinto, MJ Donnelly, CA Sousa, J. Malta-Vacas, V. Gil, C. Ferreira, V. Petrarca, VE Do Rosario, and JD Charlwood. An island within an island: genetic differentiation of *Anopheles gambiae* in Sao Tome, West Africa, and its relevance to malaria vector control. *Heredity*, 91(4):407–414, 2003. ISSN 0018-067X.
- Edward Pollak. A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104(3):531–548, Jul 1983.
- F. Simard, T. Lehmann, J.J. Lemasson, M. Diatta, and D. Fontenille. Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci. *Insect Molecular Biology*, 9(5):467–479, 2000. ISSN 1365-2583.
- L.W. Simmons. *Sperm competition and its evolutionary consequences in the insects*. Princeton Univ Pr, 2001. ISBN 0691059888.
- R R Sokal and F J Rohlf. *Biometry*. W. H. Freeman and Co.: New York, 1995. ISBN 0-7167-2411-1.
- P. Solano, S. Ravel, J. Bouyer, M. Camara, M.S. Kagbadouno, N. Dyer, L. Gardes, D. Herault, M.J. Donnelly, and T. De Meeûs. The population structure of *Glossina palpalis gambiensis* from island and continental locations in coastal Guinea. *PLoS Negl Trop Dis*, 3(3):e392, 2009.
- D.A. Tallmon, D. Gregovich, R.S. Waples, C. SCOTT BAKER, J. Jackson, B.L. Taylor, E. Archer, K.K. Martien, F.W. Allendorf, and M.K. Schwartz. When are genetic methods useful for estimating contemporary abundance and detecting population trends? *Molecular Ecology Resources*, 10(4):684–692, 2010. ISSN 1755-0998.

- C.E. TAYLOR, Y.T. TOURE, M. COLUZZI, and V. PETRARCA. Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. *Medical and veterinary entomology*, 7(4):351–357, 1993. ISSN 1365-2915.
- F. Tripet, YT Toure, CE Taylor, DE Norris, G. Dolo, and GC Lanzaro. DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Molecular Ecology*, 10(7):1725–1732, 2001. ISSN 0962-1083.
- C. Walton, J.M. Handley, et al. Population structure and population history of *Anopheles dirus* mosquitoes in Southeast Asia. *Molecular Biology and Evolution*, 17(6):962, 2000. ISSN 0737-4038.
- J. Wang. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetics Research*, 78(3):243–257, Dec 2001.
- Jinliang Wang and Michael C Whitlock. Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, 163(1):429–446, Jan 2003.
- R. S. Waples. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 121(2):379–391, Feb 1989.
- Robin S. Waples. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci\*. *Conservation Genetics*, 7(2):167–184, April 2006.
- Robin S. Waples and Chi Do. ldne: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, 8(4):753–756, 2008. URL <http://dx.doi.org/10.1111/j.1755-0998.2007.02061.x>.
- R.S. Waples and C. Do. Linkage disequilibrium estimates of contemporary  $N_e$  using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 3(3):244–262, 2010. ISSN 1752-4571.

- D. Weetman, C.S. Wilding, K. Steen, J.C. Morgan, F. Simard, M.J. Donnelly, and K.Y.K. Chan. Association Mapping of Insecticide Resistance in Wild *Anopheles gambiae* Populations: Major Variants Identified in a Low-Linkage Disequilibrium Genome. *PloS one*, 5(10):e13140, 2010. ISSN 1932-6203.
- C. Wondji, F. Simard, T. Lehmann, E. Fondjo, A. SAMÈ-EKOBO, and D. Fontenille. Impact of insecticide-treated bed nets implementation on the genetic structure of *Anopheles arabiensis* in an area of irrigated rice fields in the Sahelian region of Cameroon. *Molecular Ecology*, 14(12):3683–3693, 2005. ISSN 1365-294X.

# Tables

Species	Sampling	N. alleles	Temporal spacing	$\hat{N}_E$	Motivation	Publication
<i>A. gambiae</i>	11/55	K between 45 and 64	85, 98	4,258–6,359		Lehmann et al. [1998]
<i>A. arabiensis</i>	9/50	7	4, 9, 40	229–1046	Seasonal	Simard et al. [2000]
<i>A. gambiae</i>	12/55	NA	12, 24	1049, inf, 1457	DDT impact <sup>1</sup>	Pinto et al. [2003]
<i>A. arebiensis</i>	12/55	8	5, 12, 16, 33	135–649	ITN	Wondji et al. [2005]
<i>G. palpalis</i>	12/35	10	23	229–1046		Dyer et al. [2009]

Table 1: Representative sample of empirical studies of contemporary  $N_e$ . The sampling strategy includes approximate number of loci and individuals sampled. The number of alleles is an average approximation of the reported value (with the exception of [Lehmann et al., 1998] where the total number of independent alleles is reported). The sample spacing is in months.

## Figure legends

Figure 1: Boxplot charts of temporal point estimates for a time span between samples of 4, 12 and 24 generations. Four sampling strategies are considered: 60 individuals with 10, 20 and 50 loci and 120 individuals with 10 loci. The first line reports a constant  $N_c=500$ , the second line 1,000 and the third 2,000.

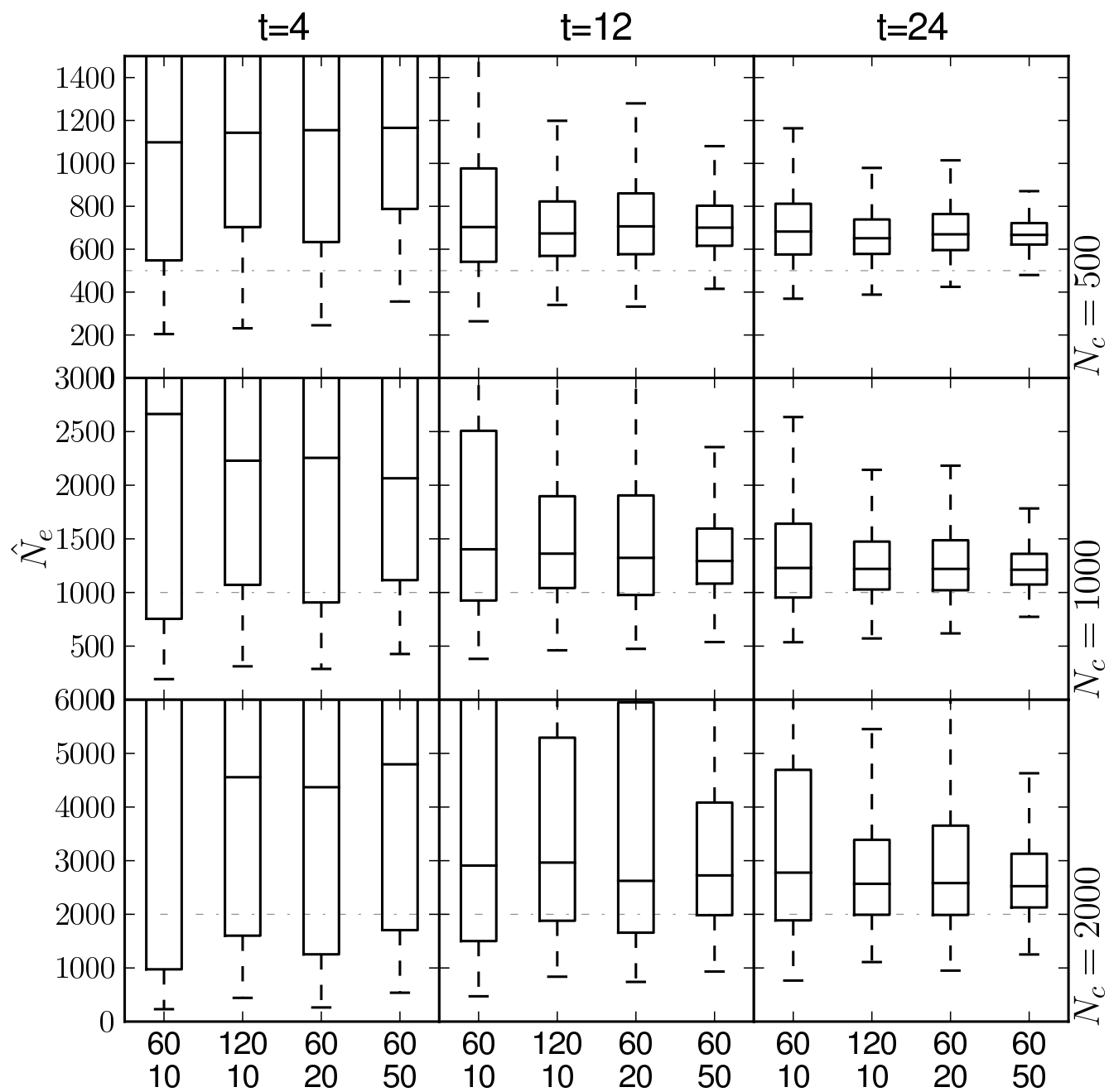
Figure 2: Harmonic mean of  $\hat{N}_E$  (solid line) and 95% confidence intervals (dashed lines) of 1,000 replicates for the moment-based temporal estimator for a time span between 1 and 24 generation. Four sampling strategies are considered: 60 individuals with 10, 20 and 50 loci and 120 individuals with 10 loci. The first line reports a constant  $N_c=500$ , the second line 1,000 and the third 2,000.

Figure 3: Boxplot charts for the point estimates of all three methods using different sampling strategies (60 individuals with 10, 20 and 50 loci) and constant  $N_c$  of 500 and 1,000. For the both temporal estimators two time spans were considered: 4 and 24 generations. The moment based version is called “NeF”, the LD version “LD” and the ML version “TM”.

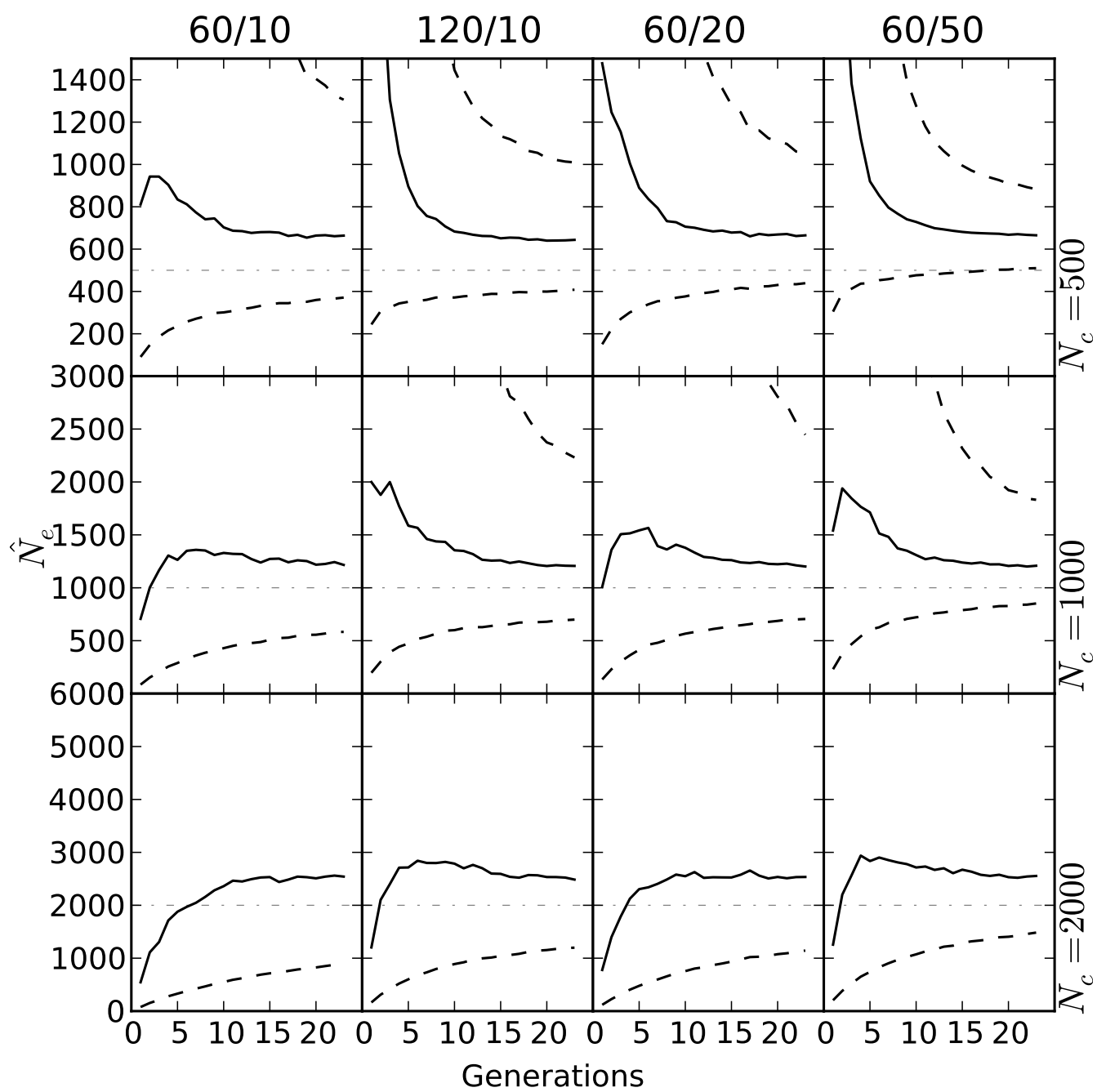
Figure 4: The behaviour of all estimators with a seasonal model. The top chart presents the harmonic mean of the point estimates for all estimators using 60 individuals and 50 loci. For the temporal methods the two reference generations were used: generation 10 with an maximum  $N_c$  of 1,000 and generation 16 with a minimum  $N_c$  of 2000. The bottom chart presents the boxplot distributions of the LD estimator.

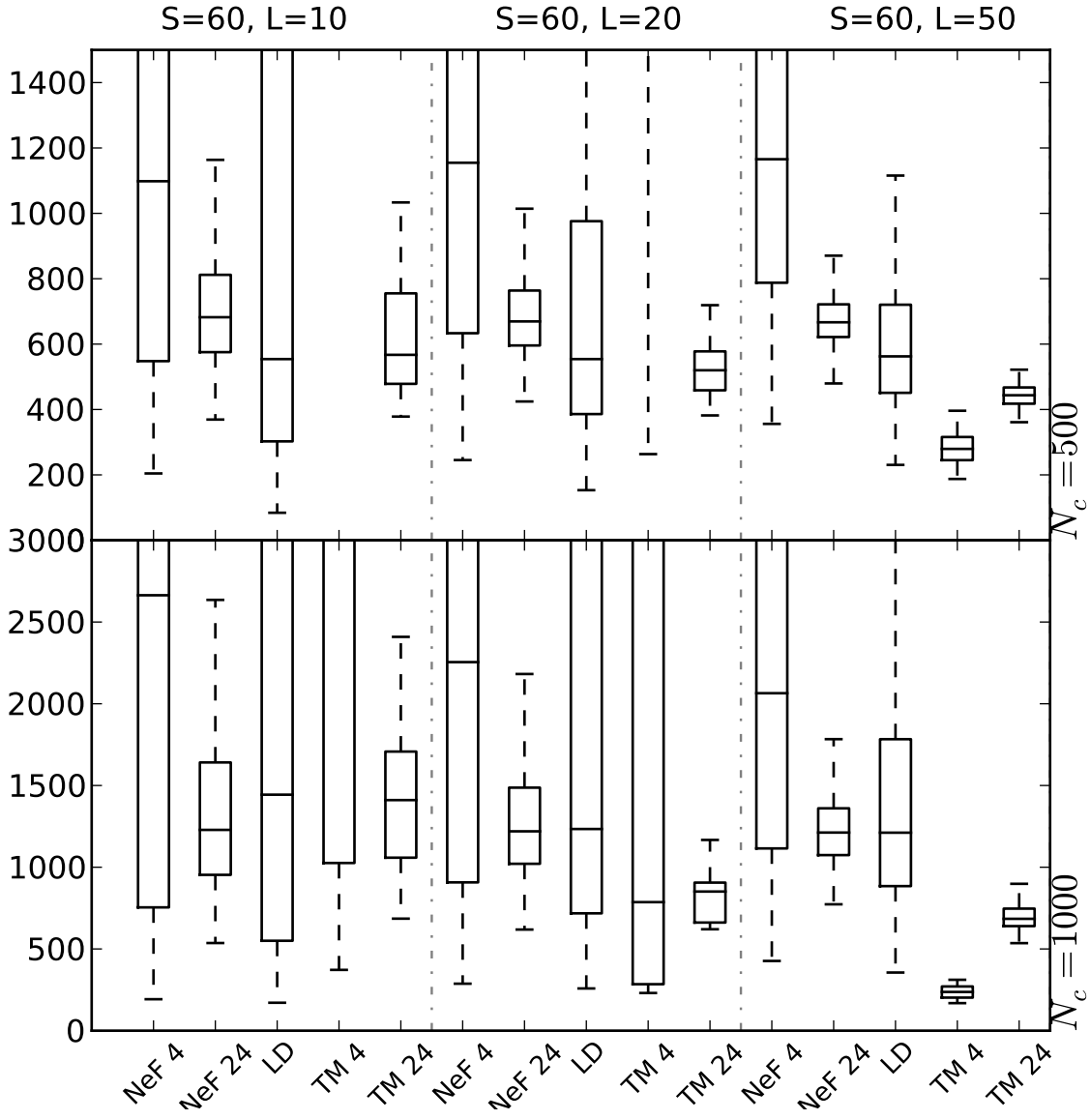
Figure 5: Harmonic mean of  $\hat{N}_E$  (solid line) for all methods methods for two bottleneck scenarios ( $N_c$  from 2,000 to 200 and from 5,000 to 500) and a sampling strategy with 60 individuals and 50 loci. Generations are counted from after the bottleneck event.

## Figures



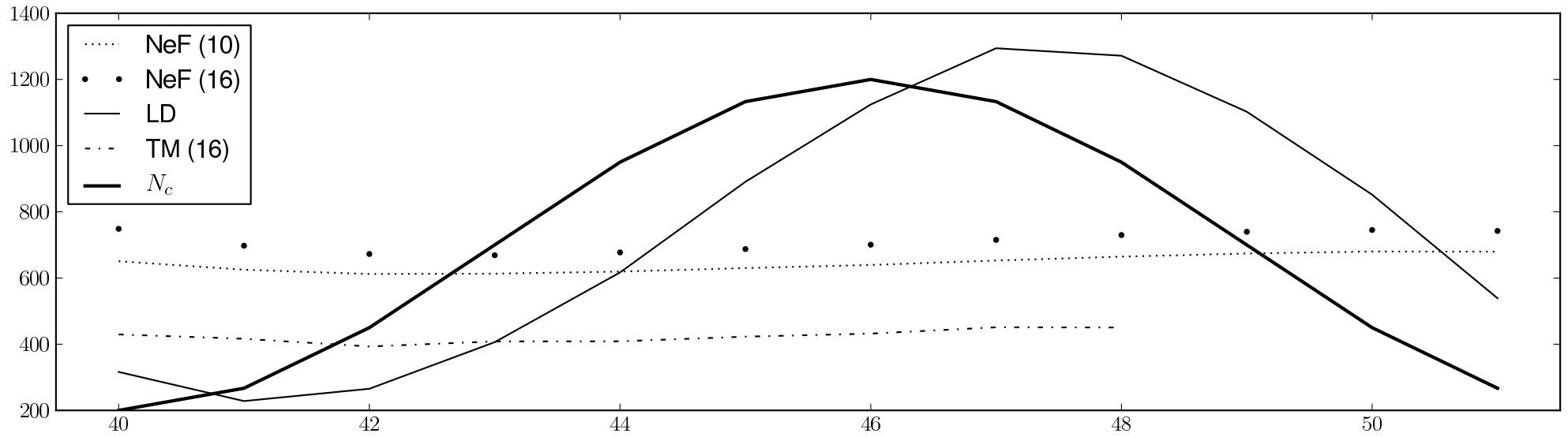






$\hat{N}_e$  of wave model

Harmonic mean of temporal and LD methods



Boxplot of point estimates (LD method)

