Estimating the number of smallmouth bass in Mille Lacs, Minnesota in 2012.

Prepared for the Minnesota Department of Natural Resources
By
Carl James Schwarz
Department of Statistics and Actuarial Science
Simon Fraser University
8888 University Drive
Burnaby, BC, Canada V5A 1S6
cschwarz@stat.sfu.ca

2014-10-17

## 1. Introduction

The Minnesota Department of Natural Resources conducted a mark-recapture experiment to estimate the number of smallmouth bass in Mille Lacs, Minnesota. Marking of fish occurred in 2012; recapture of fish in in 2013 The marking protocol is fully described in Jensen (2013) and the recapture protocol is described in Jones (2013).

Briefly, approximately 560 smallmouth bass were captured on the spawning grounds in late April and early May 2012 (Tables 1a and 1b). About 2/3 of the fish were sexed, all were fin clipped on the third dorsal spine and released without tags. Other variables (such as location of the spawning ground, length of the fish, etc.) were also recorded. Note that the tagging records are not present in the master database, but were provided in an email dated 2014-08-27.

Sampling gillnets were set between late-May to late-June in 2013. Approximately 850 unclipped/untagged smallmouth bass were captured, plus an additional 20 recaptures from the fish released on the spawning grounds as determined by the presence of the fin clip. The sex of the fish was determined for less than 10% of the new captured fish and recaptured fish. Other variables (such as location of the gillnet, length of the fish, etc.) were also recorded.

There is a very distinct difference in the selectivity in the tagging and gillnet samples (Table 1b), where the gillnets apparently did not select fish smaller than 9". For this reason, the population of interest in this report is smallmouth bass greater than 9" in length. The usual assumptions for a closed population capture-recapture study are made, including

- Closure of the population between marking and recapture. Because fish were marked in 2012 and recaptured in 2013, the assumption of closure is likely not satisfied. However, if only mortality/harvest can be assumed to have occurred between the two sampling events, the estimates of abundance are still unbiased for the population alive at the time of marking in 2012. If only recruitment to the



Consultant's Report

population occurred (e.g. from smaller length classes), then the estimates will be unbiased for the number of fish at the time of recapture. If both mortality/harvest and recruitment occurred, then estimates of abundance are unbiased for the total population ever present, i.e. fish alive at the time of tagging in 2012 (including subsequent harvest/mortality) plus new recruits.

- Marks are not lost between sampling occasions. Because fish are fin-clipped and the time interval is short, clips are unlikely to be "lost" due to regrowth.
- Marked fish can be correctly identified. The examination of fish was done by MDNR members so this again seems reasonable. Fish in 2014 were also tagged with numbered tags to differentiate the fin-clipped fish from those in 2013 (no tags) and 2014.
- Mixing of tagged and untagged fish. Tagging/clipping was done at many spawning locations around the lake as was the gillnet survey. However, there was only a short period of time between the end of the marking and the start of the gillnet sampling. If tagged/clipped fish have not fully mixed with other fish from spawning grounds not sampled, there is the potential for substantial bias.

Given the sparseness of the data, only a small number of estimation methods will be applied to this experiment. All estimation was performed using R 3.1.1 (R Core Team 2013)

## 2. Pooled-Petersen Estimator

A breakdown of the number of fish clipped, recaptured, and newly recaptured is found in Table 2. The simple Petersen estimates, combined over both sexes are

2012 23 (SE 5) thousand fish

The simple Petersen estimate can be biased because of heterogeneity in catchability between the two sexes or over covariates (such as length) or a differential change in catchability between the two sampling occasions for the two sexes. Seber (1982) showed that the bias in a simple-Petersen estimate is related to the negative correlation of the catchability between the first and second sampling occasion. For this reason, several stratified estimators were attempted.

## 3. Stratified-Petersen on the basis of sex alone

If sex was measured for all fish at both sampling occasions, then it is relatively simple to compute a stratified-Petersen estimate by finding the Petersen estimate for each sex separately, and then adding the two estimates together. Unfortunately, the sex of the fish was not determined for many of the fish (Tables 1a and 1b).

Unfortunately, the partially-stratified Petersen estimator used in the walleye analysis cannot be applied to this experiment. This method assumes that the fish that were sexed/not sexed at the time of tagging can be identified at the time of recapture (by either a different location for the mark by sex, or by unique tag numbers). Because all fish were clipped on the dorsal fin without individual tags, it is impossible to know if a fish sexed/not sexed at the time of tagging is recovered, i.e. it is impossible to determine the number of fish with histories M0, MM, F0, FF, U0, UU).

We attempted a stratified estimate by imputing sex based on the length of the fish. A logistic regression was fit to the all of the small bass data in the database where sex was recorded. The probability of being a male was modeled as a function of length (Figure 1). This fit as NOT statistically significant, i.e. there is no evidence that the length of fish is related to it sex. Nevertheless, this fitted model was used to classify an unsexed-fish as male/female using a binomial distribution depending on the predicted probability of being male. If the sex was known, the known sex was used. The imputed summary statistics are shown in Table 2.

If these imputed values are used directly in a fully stratified Petersen estimator (i.e. a separate estimate for each sex), this gives estimates of

2012 22 ( 12 F, 10 M) thousand fish with standard errors[1] of
5 ( 3 F, 4 M)

This estimate is very similar to the pooled-Petersen values. The estimated catchability on the spawning ground was similar for both males and females, again leading to the case where the Pooled-Petersen may be unbiased.

## 4. Stratification by length - I

Catchability may also varies by length within each sex at both samples because of gear selectivity. Pure heterogeneity in catchability (e.g. gear selectivity with the same shape in both samples) implies that there is a positive correlation in catchability between the two samples which could lead to a negative bias in the estimates of population abundance.

Figure 2 shows that the distribution of lengths in all the samples. The data is very sparse but does not show any major differences in the length distributions between the two sexes.

Unfortunately, unlike the analyses for walleye and northern pike, the data is too sparse and sex determination too uncertain to do a stratification by sex and length. Consequently, only a stratification by length was attempted. Because of the relatively small number of recaptures, two length strata were defined as 9-17", and 17+" to get approximately equal number of recaptures in each length class. Because of the small number recaptures (20 in total with 9 at 16"), there is no ideal split that will give sufficient numbers of recaptures in all of the strata. The summary statistics are shown in Table 3.

---

[1] These SE are slightly too optimistic because of the randomness in imputing sex has not been accounted for in 2013.

Estimates of the abundance for males and females for each stratum and the overall total were obtained (Table 4). The overall population estimates are

2012 22 (SE 5) thousand fish

The length stratified estimate in 2013 is similar to the unstratified estimates. The estimated capture rates in the two strata are very similar at the spawning sample – equal catchability among strata at any capture occasion leads to unbiased pooled-Petersen estimates.

## 5. Stratification by length - II

The previous section required that the length be divided into a small number of strata. Chen and Lloyd (2000) developed a method where no stratification is needed – a smoothed estimate over all lengths is used to estimate a Petersen estimate based on a moving window. Unfortunately, the data is too sparse and sex identification too poor to use both length and sex in the model as was done for walleye and northern pike and so this method is only applied to the combined sexes.

Summary statistics are presented in Table 5. Note that length classes were truncated at 22" because the number of fish in the length interval was too small to be useful. This may lead to a slight underestimate of abundance by ignoring fish in these higher (missing) length classes.

The estimated M+F abundance (by length) is shown in Figures 3. The estimated total population abundance using the modified Chen and Lloyd method is

2012 26 (SE 8) thousand fish

Figure 3 shows an apparent drop off in population abundance at lower lengths – this may be an artifact of the steep decline in selectivity by length. Spawning also begins in the 10-12" range so fish smaller than this would not be present on the spawning ground. Smallmouth bass seldom grow longer than 20" so there will be a mixture of many age classes in the 15-18" range.

The estimated total abundances are similar to the previous estimates. Figure 4 shows the estimated selectivities in the two surveys as a function of length. Selectivity was fairly flat on the spawning ground but very selective in the gillnet survey. This was also evident in the raw data of Table 1a. Again, equal selectivity over all fish in one of the surveys is one of the conditions under which the pooled-Petersen estimator remains unbiased.

## 6. Discussion

Estimates from the various methods were all very similar for all methods (Figure 5) with confidence limits that are quite wide because of the very small number of recaptures. There is little basis on which to choose one estimate over the other.

The key unresolved issue with these estimates is the fact that fish less than 9" in length were ignored and fish outside of 16-19" have poor selectivity in the gillnet survey.

Substantial numbers of fish <9" were captured during the spawning ground surveys (which is a bit surprising given that spawning typically begins at the 10-12" range), but the gillnets have such poor selectivity at these lengths that this population segment is essentially "invisible". If one is willing to make a very tenuous assumption that the capture rate on the spawning grounds was the same for these smaller fish as for the fish 9" or larger, than a guesstimate to the abundance in these lower lengths is found using a proportional assumption to be about 7% of the abundance of fish of 9" or greater.

Table 1b and Figure 4 shows that selectivity rapidly falls off outside the 16-19" range. This makes it very difficult to obtain good estimates of the population abundance unless very large numbers of fish in these ranges are marked so that at least some recaptures are obtained. In the absence of such recaptures, population estimates will have limited usefulness.

**References.**

Chen, S., and Lloyd, C. (2000). A non-parametric approach to the analysis of two stage mark-recapture experiments. Biometrika 88, 649–663.

Jensen, E.T. (2013). 2012 Large Lake Sampling Program Assessment Report for Mille Lacs Lake 2012.

Jones, T.S. (2013). Population estimates of walleye and northern pike in Mille Lacs Lake, Minnesota. Minnesota Department of Natural Resources, St. Paul.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rivest L-P. and Levesque T. (2001). Improved log-linear model estimators of abundance in capture-recapture experiments. Canadian Journal of Statistics, 29, 555-572.

Seber, G. A. F. (1982). Estimation of Animal Abundance, 2nd Edition.

| Table 1a. Summary statistics for the 2012-2013 survey. | | | | |
|---|---|---|---|---|
| | Female | Male | Unknown | Total |
| Gross Tagged/clipped 2012-04-19 -> 2012-05-17 | 218 | 224 | 122 | 564 |
| (fish less than 9.5") | -0 | -0 | -39 | -39 |
| Net Tagged/clipped | 218 | 224 | 83 | 525 |
| | | | | |
| Recaptured in gillnet 2013-05-23 -> 2013-06-21 | 0 | 2 | 18 | 20 |
| | | | | |
| Gross newly captured in gillnet | 39 | 31 | 782 | 852 |
| (fish less than 9.5""") | -0 | -0 | -0 | -0 |
| Net newly captured in gillnet | 39 | 31 | 782 | 852 |

Table 1b. Summary of database as provide in email dated 2014-08-27.

| Length | Male Marked | Female Marked | Unknown Marked | Total Marked | Newly Captured | Recap |
|---|---|---|---|---|---|---|
| 1 | | | | 0 | | |
| 2 | | | 1 | 1 | | |
| 3 | | | | 0 | | |
| 4 | | | 1 | 1 | | |
| 5 | | | 10 | 10 | | |
| 6 | | | 12 | 12 | | |
| 7 | | | 9 | 9 | | |
| 8 | | | 6 | 6 | | |
| 9 | | | 4 | 4 | | 1 |
| 10 | 2 | | 7 | 9 | 4 | |
| 11 | 2 | 1 | 4 | 7 | 7 | |
| 12 | 18 | 8 | 6 | 32 | 16 | |
| 13 | 28 | 14 | 4 | 46 | 18 | |
| 14 | 26 | 29 | 3 | 58 | 15 | 2 |
| 15 | 34 | 41 | 10 | 85 | 87 | |
| 16 | 35 | 36 | 13 | 84 | 278 | 9 |
| 17 | 27 | 26 | 9 | 62 | 257 | 5 |
| 18 | 20 | 21 | 9 | 50 | 109 | 3 |
| 19 | 16 | 28 | 6 | 50 | 42 | |
| 20 | 14 | 12 | 8 | 34 | 18 | |
| 21 | 2 | 2 | | 4 | 1 | |
| 22 | | | | 0 | | |
| Total | 224 | 218 | 122 | 564 | 852 | 20 |

| Table 2. Summary statistics used for the Petersen estimators. | |
|---|---|
| **Summary statistics for Pooled-Petersen estimator** | |
| | **2013** |
| $n_1$ (marked) | 525 |
| $m_2$ (recaptured) | 20 |
| $n_2$ (sample 2) | 872 |
| **Summary statistics for Stratified-Petersen estimator with imputed values for number in each sex all samples.** | |
| $n_{1f}$ (marked) | 273 |
| $n_{1m}$ (marked) | 252 |
| $m_{2f}$ (recaptured) | 7 |
| $m_{2m}$ (recaptured) | 13 |
| $n_{2f}$ (sample2) | 591 |
| $n_{2m}$ (sample2 ) | 281 |

| Table 3. Summary statistics used for stratification by length in 2013. | |
|---|---|
| | Total |
| Tagged/clipped | |
| 09-17 in | 325 |
| 17+ in | 200 |
| | |
| Recaptured in gillnet | |
| 09-17in | 11 |
| 17+ in | 9 |
| | |
| Newly captured in gillnet | |
| 09-17 in | 398 |
| 17+ in | 452 |

| Table 4. Abundance Estimates from length stratified model in 2013 (thousands) | | |
|---|---|---|
| Stratum | M+F Est | M+F SE |
| 09-17 in | 9 | 2 |
| 17+ in | 13 | 5 |
| ALL | 22 | 5 |

| Table 5. Summary data for 2013 for modified Chen and Lloyd method. Males + Females | | | | |
|---|---|---|---|---|
| Length Centre (inches | Tagged/ Clipped | Recap | Imputed Unclipped | Recap prob |
| 9.5 | 4 | 1 | 0 | 0.250 |
| 10.5 | 9 | 0 | 4 | 0.000 |
| 11.5 | 7 | 0 | 8 | 0.000 |
| 12.5 | 32 | 0 | 15 | 0.000 |
| 13.5 | 46 | 0 | 18 | 0.000 |
| 14.5 | 58 | 2 | 17 | 0.034 |
| 15.5 | 85 | 0 | 85 | 0.000 |
| 16.5 | 84 | 9 | 296 | 0.107 |
| 17.5 | 62 | 5 | 238 | 0.081 |
| 18.5 | 50 | 3 | 109 | 0.060 |
| 19.5 | 50 | 0 | 42 | 0.000 |
| 20.5 | 34 | 0 | 18 | 0.000 |
| 21.5 | 4 | 0 | 1 | 0.000 |

Figure 1. Logistic regression used to impute sex when not recorded as a function of length. The final fitted equation is logit(male) = -1.87+0.60(length), but the fit was not statistically significant. No fish larger than 20'' ever had the sex recorded.

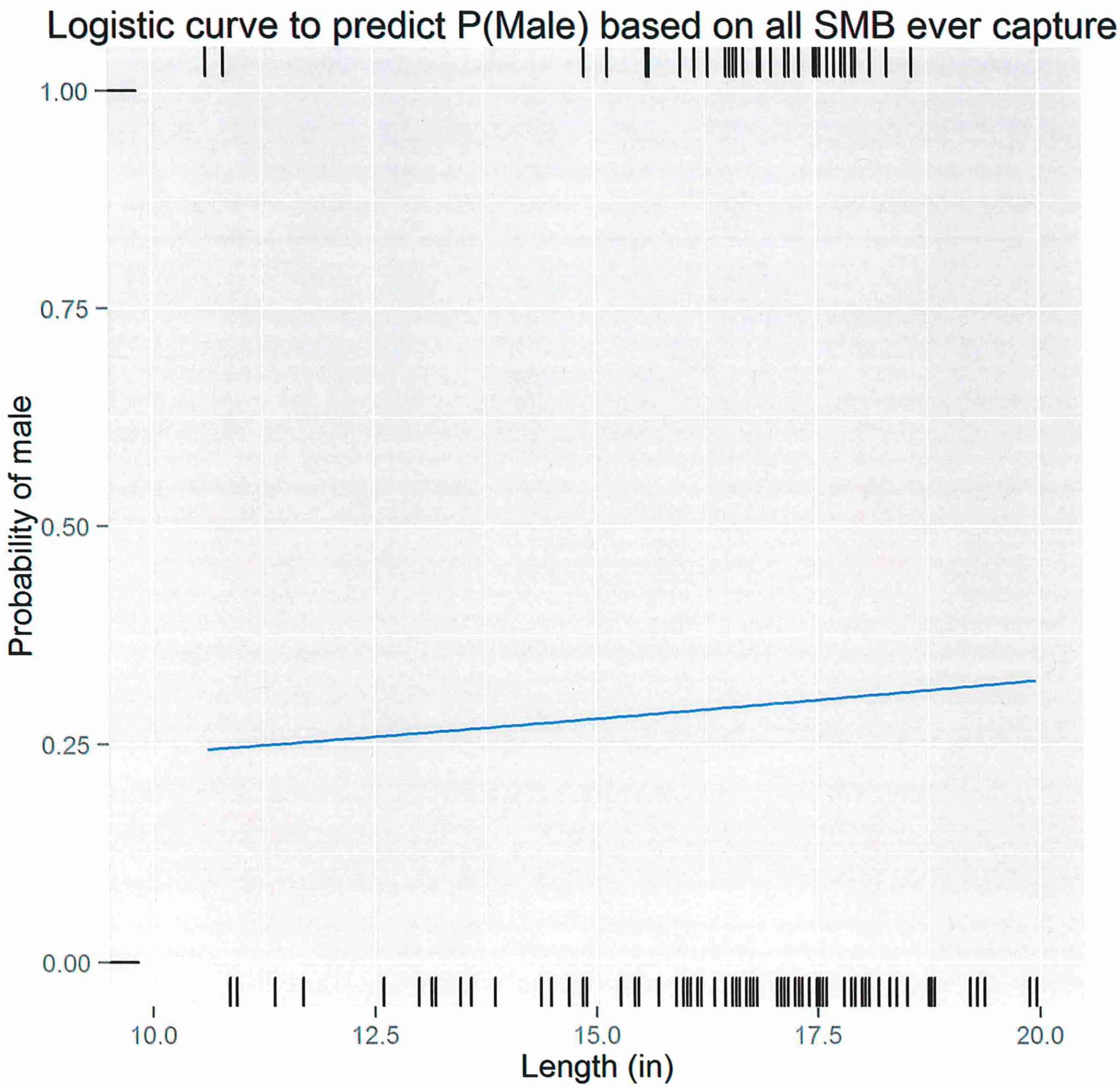

Figure 2. Summary of distribution of lengths in the various samples in 2013.

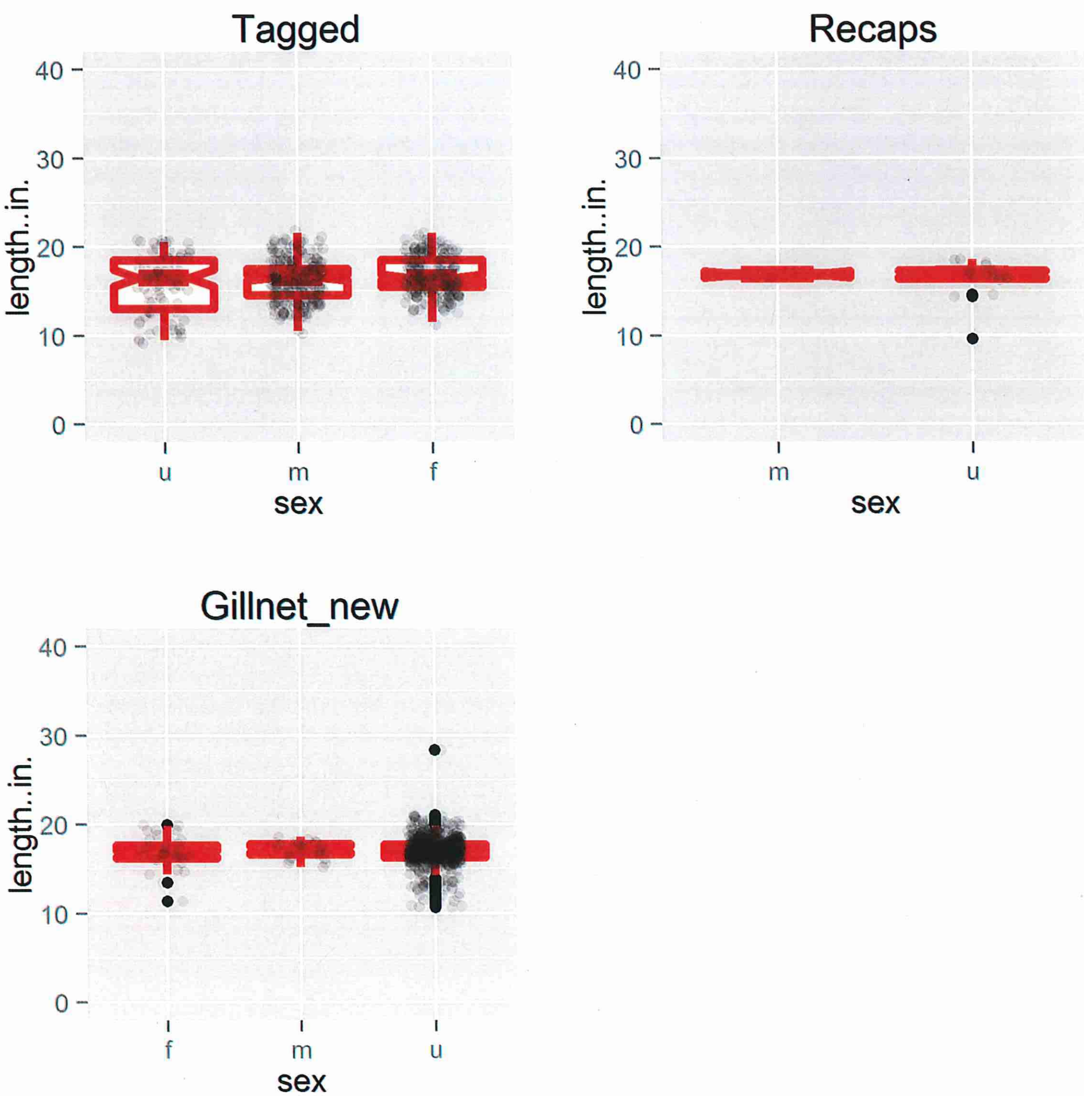

Figure 3. Estimated abundance in 2013 from the modified Chen and Lloyd method. Points are the separate estimates computed using statistics from each length interval (Table 5). The curve is the smoothed estimates. Total abundance estimates at top of figure are derived from smoothed estimates.

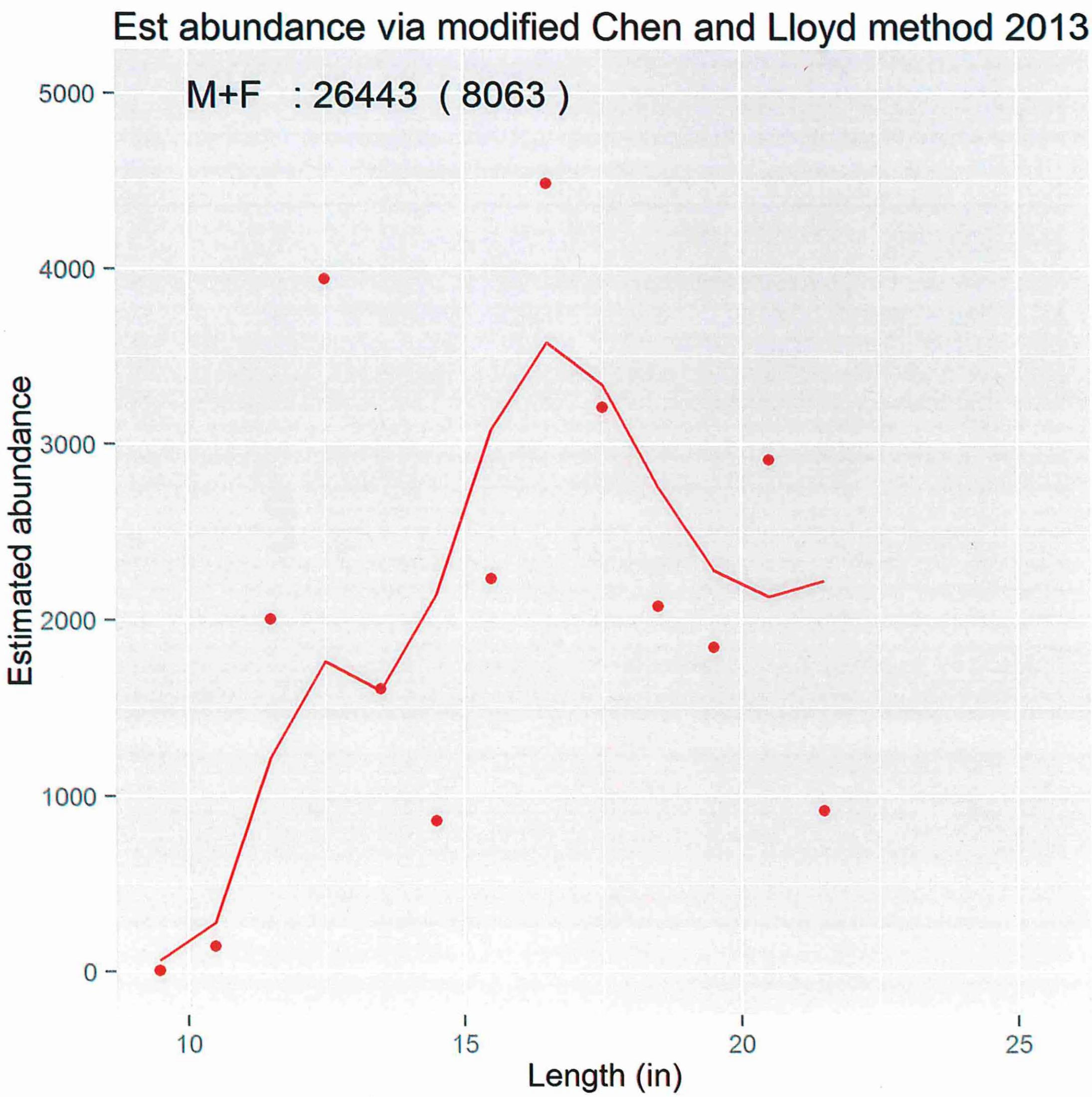

Figure 4. Estimated gear selectivity for M+F in both samples in 2013 based on the modified Chen and Lloyd (2000) method.

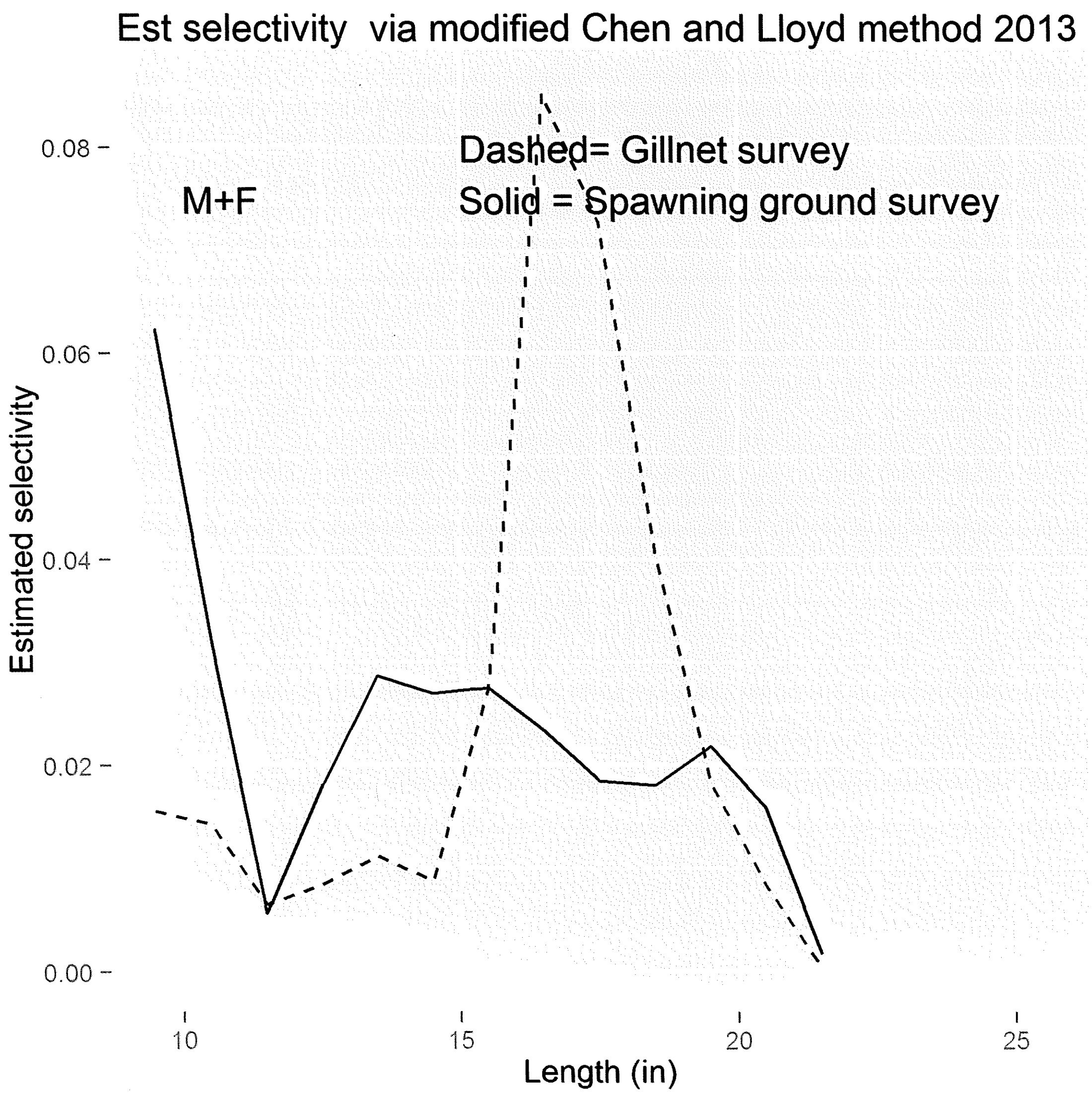

Figure 5. Comparison of estimates of total abundance from the different methods