# Estimating length composition of fish observed with stereo-video cameras: a simulation study with application to red snapper (*Lutjanus campechanus*)

Erik H. Williams[1], Kyle W. Shertzer[1], and Nate Bacheler[1]

1: Southeast Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, Beaufort, NC, USA.

## Abstract

Stereo-video cameras have become an important tool worldwide for enumerating abundance and length compositions of marine fish. The two most common approaches for enumerating are referred to as *MaxN* and *MeanCount*, where the former counts fish on the single video frame with the most individuals observed and the latter uses the mean across multiple frames. Previous studies have demonstrated that both approaches may work well for computing relative abundance. However, basic fish population age-structure (e.g. more younger fish) and fish schooling characteristics (e.g. fish of similar size/age swim together) suggest the potential for bias in length compositions computed from the *MaxN* approach, as the single video frame with the most individuals would tend to overrepresent smaller, more populous fish. To evaluate the two approaches, we simulated a stationary video system placed on a sampling site inhabited by moving fish, and we compared length compositions estimated from each approach to that of the true, underlying population. Indeed, *MaxN* appears biased over a range of potential factors, in which the bias is toward estimates of mean size that are smaller than the true values. The factors leading to biased age/length estimates from *MaxN* include 1) large fish counts in the single frame, 2) fish schooling by age/length, especially at moderate school densities, and 3) small camera viewing area relative to the site being sampled. In addition to the general evaluation, we applied the simulation framework to video parameters derived from data on red snapper (*Lutjanus campechanus*) in U.S. Atlantic waters. The results suggest that the *MaxN* length composition bias is small for red snapper due primarily to the low counts of fish on site and low levels of schooling observed.

Keywords: sampling design, simulation, stereo-video methods, length composition data, red snapper

Corresponding author: Erik H. Williams

a. *Email:* erik.williams@noaa.gov
b. *Phone:* 1-252-728-8603
c. *Fax:* 1-252-728-8789

## 1. Introduction

Sampling of fish using video has become more widespread as fishery scientists continue to seek data improvements and more accurate stock assessments. Video has the advantages of being a non-lethal sampling technique for counting fish and is generally less selective than traditional sampling gears (Morrison and Carbines 2006, Bacheler et al. 2013), but also has the disadvantage of lacking basic biological data that lethal sampling approaches have historically provided (e.g., length, age, maturity, etc.). Knowing the length and age of sampled fish is vital to many modern stock assessment and population demographic methods (Ono et al. 2015). Consequently, video collection programs are increasingly using stereo-video, which allows the estimation of fish length (Cappo et al. 2003, Cappo et al. 2004, Watson et al. 2005, Watson et al. 2010, Shortis et al. 2013, Letessier et al. 2015, Langlois et al. 2015, Langlois et al. 2020, Schramm et al. 2020). This is an important advance in video data collection, but may pose issues with video enumeration methods that were established primarily to estimate abundance rather than length composition.

The enumeration of fish on video is typically based on a single frame when the most individuals of the species of interest are observed ("*MaxN*"; Watson et al. 2010, Langlois et al. 2012; also referred to as *MinCount*, because the maximum observed on any frame is an estimator of the minimum abundance on the site). The relationship between *MaxN* and true abundance may be nonlinear, however, and methods that consider multiple frames have also been developed. For example, Schobernd et al. (2014) developed an alternative unbiased video metric called "*MeanCount,*" where fish abundance is estimated as the mean number of fish observed in a series of 41 equally spaced frames over the viewing interval (see also Bacheler and Shertzer 2020). Although there are different methods for enumerating fish on video, the measurement of fish

lengths from stereo-video has almost universally occurred at *MaxN* in order to avoid repeated measurements of the same fish (Langlois et al. 2012).

Reading stereo-video data for fish lengths versus enumerating fish counts may pose different challenges with respect to statistical properties and sampling efficiency. Past studies have shown that using *MaxN* methods for computing relative abundance work well for tracking fish populations through time under most conditions (Campbell et al. 2015), despite potential nonlinear relationships with true abundance (Schobernd et al. 2014). Ideally, pilot studies and sampling designs are conducted *a priori* to optimize efficiency in data collection (Mallet et al. 2021). However, the statistical properties of length composition estimates from *MaxN* and *MeanCount* video reading has not previously been studied.

Given that fish often school by size (Pitcher and Parrish 1993, Hoare et al. 2000) and younger fish are typically more numerous than older fish for any particular species, the *MaxN* approach may be biased when measuring fish on a single video frame with the most fish observed. On the other hand, the *MeanCount* has less potential for bias given that it is not based on an extreme value statistic, but has (to our knowledge) never been used for quantifying length distributions.

Determining the bias and precision of fish lengths from *MaxN* and *MeanCount* metrics is important for accurately parameterizing stock assessment models. Length composition data sampled from a fishery-independent video gear can provide important population dynamic information on the shape of the selectivity curves, the amount of total mortality, and cohort strength. Video samples have the potential to provide important information on smaller and younger fish, potentially serving as an independent estimate of recruitment. Unbiased video samples could also be the basis for comparing length composition data from the fisheries, resulting in improved selectivity estimates in stock assessments (Ono et al. 2015).

Here, we address potential error (e.g., bias) in stereo-video estimated length compositions by simulating a stationary video system placed on a sampling site inhabited by moving fish, and then comparing length compositions estimated from *MaxN* and *MeanCount* approaches to those of the true simulated population. We evaluate how length compositions based on the two methods are affected by local fish abundance, size/age dependent schooling behavior, and the viewing area of the video system relative to the area of sampling site. Our goals are to identify conditions under which estimated length compositions from either approach may be considered accurate and to quantify bias where it occurs. We then apply the model using parameter estimates for red snapper (*Lutjanus campechanus*) in U.S. Atlantic waters to demonstrate both how the model can be applied to actual fish stocks to explore properties of estimated length compositions for this particular species.

## 2. Methods

### *2.1 Simulated Population and Sampling Designs*

We used a simple simulation study to investigate several potential factors that lead to bias in length composition data collected from stereo-video capture systems using *MaxN* and *MeanCount* methods. We simulated a stationary video sampling system placed on a single site inhabited by moving fish. More specifically, a known age-structured population of fish ($N$) was simulated on the site with various levels of schooling properties and movement in and out of the camera field of view. We configured the simulation to approximate the sampling system used by the Southeast Reef Fish Survey (SERFS), which samples reef-associated species in Atlantic waters of the southeast United States (Bacheler et al. 2013, Schobernd et al. 2014). All simulations were coded using R (R Core Team, 2021).

The SERFS is a collaborative reef fish monitoring survey consisting of multiple vessels sampling along the southeast United States Atlantic continental shelf from Cape Hatteras to Cape Canaveral using standardized methods. Sampling occurs on hard-bottom sites between 15 and 115 m deep during daylight hours using baited chevron traps outfitted with video cameras (see Bacheler and Shertzer 2020 for more details). Since 2019, stereo-video cameras have also been attached to a subset of chevron traps to provide fish length measurements that could be used to estimate selectivity patterns of video and trap gears (e.g., Langlois et al. 2015).

To simulate the video sampling process, we first generated a fish population with ages ranging from 1 to 20 years ($y$) and an equilibrium (constant recruitment) age structure conditional on an assumed annual natural mortality rate ($M$). For a given sampling site, a fish population ($N$) was randomly generated by drawing ages from a multinomial distribution with $N$ fish and probabilities determined by the equilibrium, exponentially decaying age structure. We implicitly assumed selectivity or availability of fish to the video gear was 1.0 for all ages, although our qualitative findings should be robust to this assumption. The length of the fish ($L$, with generic units) at each age ($a$) followed a von Bertalanffy growth curve with parameters $L_\infty$, $K$, and $t_0$ (Table 1, Quinn and Deriso 1999). The height of fish ($H$) was fixed at 0.5 of $L$ so that age specific two-dimensional fish area is $L_aH_a$. Variability in size at age for individual fish was modeled using a normal distribution with an age-constant coefficient of variation ($cv$) (Table 1).

To simulate a video system for detecting fish, we made a few simplifying assumptions for ease of computation (Table 2). We assume there is no error in the measurement of fish lengths from videos. Fish were simulated in two dimensions, ignoring the distance of fish from the camera. Fish were simulated as if swimming on a 360-degree cylindrical plane around the camera. The distance of the fish from the camera is then a fixed quantity and was determined in our simulation

126 through a field of view parameter, specified as a width expressed in units of fish lengths of
127 maximum size ($L_\infty$). In this case, we used a somewhat arbitrary field of view width ($w$) of eight
128 maximum sized fish, $8L_\infty$. We also fixed the proportion of the camera view ($C_p$) to 45%, or 162
129 degrees, based on the optical-physical properties of Go-Pro camera systems currently in use by the
130 SERFS. The vertical dimension of the camera viewing system was assumed to be infinite or not
131 limiting (e.g. fish stayed in the vertical view at all times). Thus, the only way fish moved in and
132 out of camera view (detection) was through horizontal movement. Fish were assumed to be
133 detected by the camera when the mid-point of their body was within the fixed field of view. Using
134 the specified values of $C_p$ and $w$, the simulated cylindrical plane is then defined in fish length units
135 of $L_\infty$ and is $w/C_p$ wide and $w$ high. The height was arbitrarily set for visual simplicity, but has no
136 impact on the simulation results. In natural systems, fish can leave the site or view of the camera
137 in ways we did not model, but this simply has the effect of reducing the effective camera view. In
138 that sense the effective viewing field of any video camera in practice is less than its optical
139 measurements alone.

140 Fish were distributed on a sample site surrounding the camera by first specifying the
141 random x-y positioning of a box to contain the school of fish. Fish schools were assumed to be
142 structured by age class, such that all fish of a particular age class would tend to school together.
143 The density of the schools ($S_d$) was specified by a value of the number of fish ($n_a$) per unit of fish
144 area ($L_aH_a$) for each age ($a$). The value of $S_d$ was assumed constant across ages. The value of $S_d$
145 constrains the positioning of the midpoint of each individual fish such that they stay within the
146 area defined by $L_aH_a$. A value of $S_d = 1$ implies the size of the school is confined to a box with
147 dimensions $n_aL_a$ by $n_aH_a$ fish length, such that 10 fish in a school would be contained within a box
148 of size equal to 10 fish lengths by 10 fish heights. A value of $S_d = 0$ implies no schooling or the

box for the school equals the entire cylindrical plane. A value of $S_d = 2$ implies 10 fish would be constrained to a box 5 fish lengths by 5 fish heights in size. Because we are only dealing with two dimensions in this simulation, $S_d \approx 9$ would be a theoretical maximum under a tight schooling assumption of fish maintaining a unit fish length distance between each other in three dimensions (i.e. sphere packing).

Movement of individual fish was governed by a two-stage randomization process. In the first stage the box containing the school of fish was repositioned by adding a random movement step size that includes the entire cylindrical plane. In the second stage of the randomization process the individual fish were repositioned with an additive random step size constrained by the box for the fish school. Random uniform numbers were drawn in both the $x$ and $y$ dimensions of the entire cylindrical plane to determine the movement step size for the fish school box positioning. A wrapping function was used to force fish whose random movement went outside the right or left side to follow a 360-degree cylinder. This randomization process was repeated at each time step, which is arbitrary, but assumes fish position is independent with each time step.

Autocorrelation in movement of the fish school from one time step (frame) to the next time step was controlled through a correlation parameter ($\rho$), specified as a value on the range [0-1]. Mathematically this was accomplished by multiplying the random movement step size by 1-$\rho$. A value of $\rho = 1$ forces no school movement in each time step, while a value of $\rho = 0$ allows for full movement anywhere within the cylindrical plane. In essence, the effect of this parameter is similar to reduced sample sizes or shortened periods of video time. If fish did not move at all, then all frames in the video would be identical and the same as a single snapshot. Thus, the movement algorithm is not explicitly connected to the speed of swimming, but rather is controlled by the

correlation parameter. A visual example of a simulated video frame with fish on a site and viewing area of the camera is shown in Figure 1.

In summary, the video simulation generates $N$ multinomial sampled random fish based on the equilibrium age structure from a known population. These fish are then randomly moved about the cylindrical plane of view for 41 video frames (time steps). All the frames together represent a *MeanCount* sample and the frame with the highest count represents the *MaxN* sample. Simulation settings are systematically changed to experimental values shown in Table 2 and compared to the 'true' population values.

### *2.2 Length Composition Data*

Using the video simulation system described above, we implemented two types of fish counting methods to derive data sets, *MeanCount* and *MaxN*. These video sampling methods rely on counting the fish in multiple video frames during a soak interval of time. To match the SERFS methods we simulated 41 frames for counting fish, using all frames for *MeanCount* samples and the single frame with the maximum number of fish for *MaxN* samples.

To evaluate performance of the estimators, we created *MeanCount* and *MaxN* samples of lengths and compared these estimates to the known lengths of the site's fish population. More specifically, we compared mean length estimates and length distributions (compositions) between the control (true) population and the simulated video samples. To compute compositions for the *MeanCount* estimator, length samples from the 41 frames were pooled together; for the *MaxN* estimator and the true composition, lengths were simply enumerated from the single frame or population, respectively. Length measures were averaged to produce mean length estimates and measures were binned into count vectors of 19 equal sized bins from fish sizes of 0 to $L_\infty$ (1000)

194 plus a bin for all fish over $L_\infty$ to produce proportional length compositions (see length data in
195 Figure 2).

196 Mean lengths were examined as the distribution of the differences between sampled and
197 simulated (true) mean lengths using the R package 'vioplot' (Adler and Kelly, 2021). Difference
198 distributions with means not equal to 0 is an indicator of bias and the spread of the distributions is
199 an indicator of variance. We also report the proportion of runs (p-value) that were found to have
200 statistically significant similar length distributions (multinomial vectors) using the R package
201 'XNomial' (Engels 2015). This is considered an exact test based on a randomization process to
202 evaluate whether the multinomial vectors of true length composition and sampled length
203 composition are similar. In this study, we applied 100,000 randomization trials with the function
204 'xmonte.' The primary advantage is that it avoids potential issues of asymptotic approximations
205 (e.g. likelihood ratio or chi-square), but comes at the cost of extra computation time.

206

207 *2.3 U.S. South Atlantic Red Snapper Case Study*

208 The video simulation described above depends on multiple parameters that are likely to vary
209 temporally during sampling, as well as by the size of the population and schooling nature of the
210 species being studied. As a case study, we used stereo-video data collected by SERFS on red
211 snapper (*Lutjanus campechanus*). The data were collected from 2018 to 2019 with a sample size
212 of 72 videos that were read for lengths following the *MeanCount* protocols. The data included
213 measurements of individual fish sizes, as well as x-y-z positions of fish within a video frame. We
214 used this data set to estimate the range and statistical distribution of $n$ fish per frame and schooling
215 density ($S_d$) based on the red snapper positions in the video frame. For simplicity we did not attempt
216 to break the red snapper data into size categories for determining $S_d$, rather $S_d$ was computed for

all red snapper in the frame. Exponential, negative binomial, gamma, and lognormal statistical distributions were fit to the values of $n$ and $S_d$ measured for red snapper in each video frame. The best fit distributions were then used to draw random values for $n$ and $S_d$ in the simulation analysis described above, which was run for 20,000 trials.

## 3. Results

### *3.1 General results*

Applying the fixed values of $N = 40$, $M = 0.2$, $S_d = 1$, $\rho = 0$, and $C_p = 0.45$ produced a clear bias in the estimated age and length composition from the *MaxN* method (Figure 2). The multi-modal pattern shown in the length composition was the result of modeling ages 1 and 2 with $cv = 0.2$, resulting in distinct age classes showing in the length composition data. The pattern of positive bias at the smaller size/ages and negative bias at the larger size/ages is a consistent result in our study when bias in the *MaxN* method is present. This bias pattern results in smaller mean length in the sample relative to the true value in every case. Therefore, a simpler representation of this pattern is a measure of difference in mean length from the sample and true population. In contrast, the *MeanCount* method indicated no bias in the composition estimates.

The effect of the number of fish on site was tested using a range of $N = 5$ to 200, with fixed parameters of $M = 0.2$, $S_d = 1$, $\rho = 0$, and $C_p = 0.45$. These simulations suggest the *MaxN* estimates are unbiased at the lowest number of fish on site ($N = 5$), but the negative bias in mean length increases with increasing number of fish on site. The increase in bias appears to asymptote at the highest levels of $N$ fish on site. The *MeanCount* estimates appear to be unbiased for all values of $N$ (Figure 3). The variance in mean length differences for *MaxN* and *MeanCount* decreases with increasing $N$, but is notably larger for the *MaxN* estimates overall (Figure 3). At the lowest value

240 of $N = 5$, the proportion of length composition estimates that match the true value is much higher
241 for *MaxN* compared to *MeanCount*, but for $N \geq 20$, the *MeanCount* length composition estimates
242 match the true values much better than the *MaxN* estimates. At the highest value of $N = 200$, the
243 *MaxN* length composition estimates become so biased that only 2% of estimates statistically match
244 the true length compositions (Figure 3).

245 Next, we tested the effects of changes in $M$ using a range of 0 to 0.5 ($y^{-1}$). In reality, $M = 0$
246 is impossible, but in this scenario, we were essentially testing the magnitude of the differences
247 between younger (smaller) and older (larger) fish as increasing $M$ would tend to skew the
248 population toward younger individuals. *MaxN* estimates showed a pattern of increasing bias with
249 increasing $M$ values up to $M = 0.2$ and then slightly decreasing bias out to values of $M = 0.5$ (Figure
250 4). Also, the proportion of length composition estimates from *MaxN* that matched the true values
251 increased steadily over the range of $M$ values. *MeanCount* estimates remained unbiased, had a
252 lower variance in length differences, and had very good matches in the length composition
253 estimates over the full range of $M$ values (Figure 4).

254 This simulation analysis was initially motivated by the suspected potential for bias of
255 age/length composition estimates resulting from fish schooling ($S_d$). Both *MaxN* and *MeanCount*
256 estimates appear to be unbiased when there is no schooling effect in place (Figure 5). As the
257 schooling density increases, the simulations indicate that the bias in *MaxN* estimates increases
258 quickly, reaching a saturation level at $S_d \geq 0.25$. For *MeanCount* the bias was near zero across the
259 range of schooling densities, but variance increased with higher schooling density. The length
260 composition estimates from the *MaxN* method showed an improved match percentage with
261 increasing values of $S_d$, while the *MeanCount* estimates seem to match almost perfectly across all
262 values of $S_d$ (Figure 5). The case of $S_d = 0$ suggests that even when fish schooling is not present,

263 the *MaxN* method is unable to match the true length composition, with only 52% of the estimates
264 indicating statistically significant matches (Figure 5).

265 We next evaluated the effects of various levels of correlation ($\rho$) in movement on both
266 *MaxN* and *MeanCount* estimates. Reduced movement, as indicated by higher $\rho$ values, does not
267 start to affect length estimates until it reaches extreme values of $\rho = 0.75$ and 1.0 (Figure 6). At the
268 value of $\rho = 1.0$, which implies no fish movement, neither *MaxN* nor *MeanCount* are able to
269 estimate the length composition reliably (Figure 6).

270 Last, we evaluated the effects of the camera view ($C_p$) on length composition estimates
271 from the two reading methods. Not surprisingly, smaller $C_p$ values increased variance in both
272 *MeanCount* and *MaxN* samples, while increasing values approached complete sampling (a census)
273 and thus perfect estimation (Figure 7). There is zero bias for the *MaxN* estimates at $C_p = 0.95$, as
274 indicated by an absent violin plot and a matching statistic of 1.0 (Figure 7). It should be noted that
275 even with a $C_p = 1.0$, suggesting a 360° camera system, the effective view is likely to be something
276 less than 100% due to factors that affect detection, such as turbidity.

277

278 *3.1 Red Snapper Results*

279 Red snapper stereo-video data indicated considerable variability in the number of fish per frame
280 ($n$) and $S_d$. Values for $n$ ranged from 1 to 42 and values for $S_d$ ranged from 0.2 to 4.5. The best fit
281 for $n$ came from an exponential distribution with a rate parameter equal to 0.242. To convert this
282 to a value for $N$ for use in our video simulation, we divided it by the proportion of viewing area
283 covered by the video camera, $C_p = 0.45$. This distribution suggests that we should most frequently
284 expect n<5 fish per frame, as was observed in the data. The best fit for the $S_d$ measure came from
285 a lognormal distribution with mean equal to -0.0086 and standard deviation equal to 0.5176. This

286 distribution suggests a generally low schooling value, with a mean even lower than the value of $S_d$
287 = 1 used in most of the simulation analyses above. The empirical estimates for red snapper were
288 used in a stochastic simulation, drawing randomly from the fitted distributions described above
289 using our video simulator for red snapper.

290 In the red snapper simulations, bias was observed in the *MaxN* estimates for $n > 8$, while
291 the *MeanCount* method estimates remained unbiased for all values of $n$ (Figure 8). However, more
292 than 65% of the red snapper video samples had fewer than 10 fish per frame. Thus, when all
293 simulated samples are pooled, the differences between age and length composition estimates from
294 *MaxN* and *MeanCount* nearly disappeared (Figure 9).

295

296 **4. Discussion**

297 Our simulations showed that *MeanCount* produces unbiased estimates of length
298 composition data, and compared to *MaxN*, generally has a lower variance in mean length estimates.
299 *MeanCount* estimates also tended to statistically match the true length composition data more
300 frequently than *MaxN* estimates across a wide range of conditions. *MaxN* can produce biased
301 length compositions, depending on conditions. The effect is most evident for species that show
302 strong age-dependent schooling behavior and, in our simulation, was always in the direction of
303 underestimating mean length. However, our case study for red snapper indicated very little
304 difference between length composition estimates from the *MaxN* and *MeanCount* methods, owing
305 to the low values for $S_d$ and $n$. It should be noted that field observations typically report loosely
306 aggregated schooling behavior, which could explain a dampened bias effect for that species.

307 *MaxN* is the predominant method being used around the world to count and measure fish
308 from video data (Ellis and DeMartini 1995, Merritt et al. 2011, Cappo et al. 2004, Campbell et al.

309 2018). The primary reason for this lies in the efficiency of reading one video frame for lengths
310 versus many frames for *MeanCount*. Our study suggests the optimal method for reading fish videos
311 for length data may come down to a bias versus cost (manpower) trade-off. Our simulation shows
312 that *MaxN* is unbiased when there is no fish schooling by age or size and when the numbers of fish
313 per frame are low ($n < 10$). However, when fish school by age or length ($S_d \geq 0.25$) and the number
314 of fish per frame increases ($n > 10$), then *MeanCount* will produce unbiased and statistically better
315 length composition data than *MaxN*.

316 Our case study for red snapper indicates only minor differences in length composition
317 estimates from *MaxN* and *MeanCount* methodologies, because this species exhibits relatively low
318 $S_d$ and $n$ values. The differences are so small that it seems unlikely that this would affect stock
319 assessment or population estimates relying on length composition data. We chose red snapper for
320 convenience and data availability, but it is a large predator on U.S. South Atlantic reef sites and
321 therefore may have lower site abundance and exhibit less schooling behavior than other species in
322 the region, such as vermilion snapper (*Rhomboplites aurorubens*) or greater amberjack (*Seriola*
323 *dumerili*). More stereo-video analysis is needed for other fish species to indicate typical values of
324 $S_d$ and $n$.

325 In our simulation we focused on a video camera viewing proportion of 45%, based on
326 current Go-Pro camera systems in use for SERFS. This value is likely an overestimate because of
327 factors such as water clarity, vertical movements of fish out of range of camera (ignored in this
328 simulation), and obscuring by overlapping fish, all of which will tend to reduce the effective
329 viewing area of a camera. The simulation results here indicate lower view proportions will increase
330 the bias induced by *MaxN* sampling when fish schooling is present. Our simulation also showed
331 that the bias in length composition data from the *MaxN* method diminishes at high values of camera

332 viewing ($C_p$ > 0.75). This implies a nearly 360° camera viewing system may help overcome
333 potential bias in length composition data from *MaxN*. Recent work has also shown that 360°
334 cameras improve relative abundance estimation based on *MaxN* as well (Kilfoil et al. 2017,
335 Campbell et al. 2018). This is another example of a potential bias versus cost trade-off in fish
336 length composition data collection. The cost of a 360° stereo-video system could be quite high, but
337 when balanced with the difference in cost between reading video data using *MaxN* versus
338 *MeanCount*, the cost might be justified.

339 The biases in length composition estimates from *MaxN* appear over a range of potential
340 conditions. The bias is in one direction and is driven by the differential age or length composition
341 produced by a cohort-structured population, combined with fish schooling by age or length.
342 Natural fish age and length structures rarely match the equilibrium patterns shown in Figure 2 at
343 any particular time, but the average across years will generally tend toward an equilibrium. Thus,
344 the potential bias seen in this simulation will also follow the age/length structure fluctuations
345 created by varying recruitment and mortality processes, being high in some years and lower in
346 others. The pattern of bias illustrated in Figure 2 suggests that high recruitment will likely amplify
347 the bias at younger ages, and then as that strong cohort decays, the bias will diminish. Fish
348 schooling or shoaling are highly dynamic processes that are likely habitat- and species-dependent.
349 In general, fish do tend to aggregate by similar length (Hoare et al. 2000, Pavlov and Kasumyan
350 2000). For some species, the tendency for schooling may dissipate with age. Our simulation
351 assumed the schooling density was constant across all ages and only the number fish declined with
352 age. However, schooling density is but one factor that affects sampling properties of video systems.
353 The number of fish on a particular site, the field of view of the camera, and the conditions of the
354 sampling site (e.g., currents, predators, bait) all have the potential to affect how fish schooling

355 behavior is expressed for a given video sample. Given the values for parameters from our
356 simulation that produced biased length composition estimates, the type of species most vulnerable
357 to bias would be fishes lower on the trophic scale, primarily forage fishes and other fishes known
358 to school together (e.g. sparids, scombrids, carangids). The species less likely to be affected would
359 be top predatory fish that are more solitary. This suggests that understanding schooling behavior
360 of the target species is required to determine the optimal method for video data analysis. Without
361 that understanding, *MeanCount* estimates may be the best default approach, given that it was less
362 biased over a broad range of conditions.

363 Our conclusions from this analysis are predicated on the simulation's accuracy at
364 mimicking the natural system. Regardless of the many reasons why this simulation may not exactly
365 reflect real fish swimming near a video camera, the range and number of factors we evaluated
366 support the conclusion that there is potential for the *MaxN* sampling process to produce biased
367 length composition data and that care should be taken when deciding what video reading method
368 is applied to a given species.

369 Many video sampling systems are limited to clear water in topical and sub-tropical areas
370 of the world. These systems often coincide with higher fish species diversity, thus making the
371 decision about the overall best data collection system for a multi-species system very difficult. As
372 video data collection systems shift toward stereo-video usage in fish sampling, the ability to
373 measure the depth of viewing or $z$ dimension opens up many possibilities, including possible
374 density or absolute abundance measures. Fully understanding the potential biases of *MeanCount*
375 and *MaxN* methods will be important if fish video sampling heads in that direction. In this paper,
376 we mimicked the 41 frame counts used by SERFS for the *MeanCount* method, however the
377 simulation framework could be easily modified to test the cost-benefit of different numbers of

frames to use for *MeanCount* or *MaxN*. Until we improve computer automated video reading systems, the cost trade-offs that have to be considered for the best overall data collection systems are further complicated by video processing costs. With proper accounting of video processing costs, such an analysis could point to the optimal number of frames and most efficient video sampling method overall, be it *MeanCount* or *MaxN* (Bacheler and Shertzer 2015). This study and others that may follow will hopefully aid in that decision-making process.

In our red snapper case study, the conclusion was that *MaxN* estimates would be equally valid as *MeanCount* estimates for use in demographic analysis or stock assessments. This may be the case for many other species as well. However, should a species have relatively high abundance and exhibit moderate to high levels of schooling, then *MeanCount* may provide more reliable length composition estimates. In general, length composition data in stock assessments can be important, and in particular video methods have the potential to fill important data gaps, including information on the smallest individuals (Ono et al., 2015). If sampling systems shift away from collection of physical specimens, and thus reduced age sampling, and toward video methods, the importance of accurate length data in stock assessments will only increase.

**Acknowledgements**

We thank Matt Campbell, Kevin Craig and anonymous reviewers for critical comments on previous versions of this manuscript. Mention of trade names or commercial companies is for identification purposes only and does not imply endorsement by the National Marine Fisheries Service, NOAA. The scientific results and conclusions, as well as any views and opinions

expressed herein, are those of the authors and do not necessarily reflect those of any government agency.

## References


Adler, D., Kelly, S.T., 2021. vioplot: violin plot. R package version 0.3.7 https://github.com/TomKellyGenetics/vioplot

Bacheler, N.M., Schobernd, C.M., Schobernd, Z.H., Mitchell, W.A., Berrane, D.J., Kellison, G.T., Reichert, M.J.M. 2013. Comparison of trap and underwater video gears for indexing reef fish presence and abundance in the southeast United States. Fish Res 143: 81–88. https://doi.org/10.1016/j.fishres.2013.01.013.

Bacheler N.M., Shertzer, K.W., 2015. Estimating relative abundance and species richness from video surveys of reef fishes. Fish Bull 113:15-26. https://doi.org/10.7755/FB.113.1.2

Bacheler N.M., Shertzer, K.W., 2020. Catchability of reef fish species in traps is strongly affected by water temperature and substrate. Mar Ecol Prog Ser 642:179-190. https://doi.org/10.3354/meps13337

Campbell, M.D., Pollack, A.G., Gledhill, C.T., Switzer, T.S., DeVries, D.A. 2015. Comparison of relative abundance indices calculated from two methods of generating video count data. Fish Res 170: 125–133. https://doi.org/10.1016/j.fishres.2015.05.011.

Campbell, M.D., Salisbury, J., Caillouet, R., Driggers, W.B., Kilfoil, J. 2018. Camera field-of-video and fish abundance estimation: a comparison of individual-based model output and empirical data. J Exp Mar Biol Ecol 501: 46–53.

Cappo M., Larsen, D.P., Hughes, R.M., 2004. Comparison of baited remote underwater video stations (BRUVS) and prawn (shrimp) trawls for assessments of fish biodiversity in inter reefal areas of the Great Barrier Reef Marine Park. J Exp Mar Biol Ecol 302:123–152. https://doi.org/10.1016/j.jembe.2003.10.006.

Cappo M., Harvey, E., Malcom, H., Speare, P., 2003. Potential of Video Techniques to Monitor Diversity, Abundance and Size of Fish in Studies of Marine Protected Areas, North Beach, Western Australia., pp. 455–464.

Ellis, D.M., DeMartini, E.E. 1995. Evaluation of a video camera technique for indexing abundances of juvenile pink snapper, *Pristipomoides filamentosus*, and other Hawaiian insular shelf fishes. Fish Bull 93: 67–77.

Engels, B., 2015. XNomial: Exact Goodness-of-Fit Test for Multinomial Data with Fixed Probabilities. R package version 1.0.4. https://CRAN.R-project.org/package=XNomial

Hoare, D.J., Krause, J., Peuhkuri, N., Godin, J.-G.J. 2000. Body size and shoaling in fish. J Fish Biol 57: 1351–1366. https://doi.org/10.1111/j.1095-8649.2000.tb02217.x.

Kilfoil, J.P., Wirsing, A.J., Campbell, M.D., Kiszka, J.J., Gastrich, K.R., Heithaus, M.R., Zhang, Y., Bond, M.E. 2017. Baited remote underwater video surveys undercount sharks at high densities: insights from full-spherical camera technologies. Mar Ecol Prog Ser 585: 113–121.

Langlois, T. J., Fitzpatrick, B. R., Fairclough, D. V., Wakefield, C. B., Hesp,S. A., McLean, D. L., Harvey, E.S., Meeuwig, J. J., 2012. Similarities between line fishing and baited stereo-video estimations of length-frequency: Novel application of kernel density estimates. PLoS ONE 7(11):e45973. https://doi.org/10.1371/journal.pone.0045973

Langlois, T.J., Newman, S.J., Cappo, M., Harvey, E.S., Rome, B.M., Skepper, C.L., Wakefield, C.B., 2015. Length selectivity of commercial fish traps assessed from in situ comparisons with stereo video: Is there evidence of sampling bias? Fish Res 161:145–155. https://doi.org/10.1016/j.fishres.2014.06.008

445 Langlois, T., Goetze, J., Bond, T., et al., 2020. A field and video annotation guide for baited
446 remote underwater stereo-video surveys of demersal fish assemblages. Methods Ecol Evol.
447 11: 1401– 1409. https://doi.org/10.1111/2041-210X.13470.
448 Letessier, T.B., Juhel, J-B., Vigliola, L., Meeuwig, J.J., 2015. Low-cost small action cameras in
449 stereo generates accurate underwater measurements of fish. J Exp Mar Biol Ecol. 466:120-
450 126.
451 Mallet, D., Olivry, M., Ighiouer, S., Kulbicki, M., Wantiez, L., 2021. Nondestructive monitoring
452 of soft bottom fish and habitats using a standardized, remote and unbaited 360° video
453 sampling method. Fishes 6(50):1-17. https://doi.org/10.3390/fishes6040050
454 Merritt, D., Donovan, M.K., Kelley, C., Waterhouse, L., Parke, M., Wong, K., Drazen, J.C.
455 2011. BotCam: a baited camera system for nonextractive monitoring of bottomfish species.
456 Fish Bull 109: 56–67.
457 Morrison, M., Carbines, G. 2006. Estimating the abundance and size structure of an estuarine
458 population of the sparid *Pagrus auratus*, using a towed camera during nocturnal periods of
459 inactivity, and comparisons with conventional sampling techniques. Fish Res 82: 150–161.
460 https://doi.org/10.1016/j.fishres.2006.06.024.
461 Ono, K., Licandeo, R., Muradian, M.L., et al., 2015. The importance of length and age
462 composition data in statistical age-structured models for marine species. ICES J Mar Sci
463 72:31-43. https://doi.org/10.1093/icesjms/fsu007
464 Pavlov, D.S., Kasumyan, A.O., 2000. Patterns and mechanisms of schooling behavior in fish: a
465 review. J. Icthy 40(2):S163-S231.
466 Pitcher, T.J., Parrish, J.K., 1993. Functions of shoaling behavior in teleosts. Pages 122-147 in
467 (Krebs, J.R., Davies, N.B., eds) Behavioral Ecology: An Evolutionary Approach. Oxford:

Blackwell Scientific Publications.

Quinn, T.J., Deriso, R.B., 1999. Quantitative fish dynamics. Oxford University Press, Oxford. 560 p.

R Core Team., 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schobernd, Z. H., Bacheler, N. M., Conn, P. B., 2014. Examining the utility of alternative video monitoring metrics for indexing reef fish abundance. Can J Fish Aq Sci. 71:464–471. https://doi.org/10.1139/cjfas-2013-0086

Schramm, K.D., Marnane, M.J., Elsdon, T.S., Jones, C., Saunders, B.J., Goetze, J.S., Driessen, D., Fullwood, L.A.F., Harvey, E.S., 2020. A comparison of stereo-BRUVs and stereo-ROV techniques for sampling shallow water fish communities on and off pipelines. Mar Env Res 162:105198.

Shortis, M.R., Ravanbakhsh, M., Shafait, F., Harvey, E.S., Mian, A., Seager, J.W., Culverhouse, P.F., Cline, D.E., Edgington, D.R., 2013. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. Videometrics, Range Imaging, and Applications XII, SPIE. 8791. 87910G. 10.1117/12.2020941.

Watson, D.L., Harvey, E.S., Anderson, M.J., Kendrick, G.A., 2005. A comparison of temperate reef fish assemblages recorded by three underwater stereo-video techniques. Mar Biol. 148:415-425.

Watson, D.L., Harvey, E.S., Fitzpatrick, B.M., Langlois, T.J., Shedrawi, G., 2010. Assessing reef fish assemblage structure: how different stereo-video techniques compare? Mar Biol. 157:1237-1250.

490 Table 1. Values of the biological parameters used in the population simulation.
491

| **Biological factor** | **Parameter** | **Values** |
|---|---|---|
| Age range ($y$) | $a$ | 1 to 20 |
| Natural mortality ($y^{-1}$) | $M$ | 0.2 |
| Asymptotic length (length units) | $L_\infty$ | 1000 |
| Growth rate ($y^{-1}$) | $K$ | 0.25 |
| Spawning adjustment ($y$) | $t_0$ | -0.5 |
| Height of fish (length units) | $H$ | 0.5$L$ |
| Variability in length-at-age | $cv$ | 0.2 |

492

493 Table 2. Factors used in the simulation tests which were run 20,000 times.
494

| Video Factor | Parameter | Base Value | Experimental Range |
|---|---|---|---|
| Video width of view | $w$ | $8L_\infty$ | fixed |
| Number of fish on site | $N$ | 40 | {5, 20, 40, 100, 200} |
| Natural mortality | $M$ | 0.2 | {0, 0.1, 0.2, 0.35, 0.5} |
| Schooling density | $S_d$ | 1 | {0, 0.25, 0.5, 1, 2} |
| Movement correlation | $\rho$ | 0 | {0, 0.25, 0.5, 0.75, 1} |
| Proportion of camera view | $C_p$ | 0.45 | {0.15, 0.3, 0.45, 0.65, 0.95} |

495

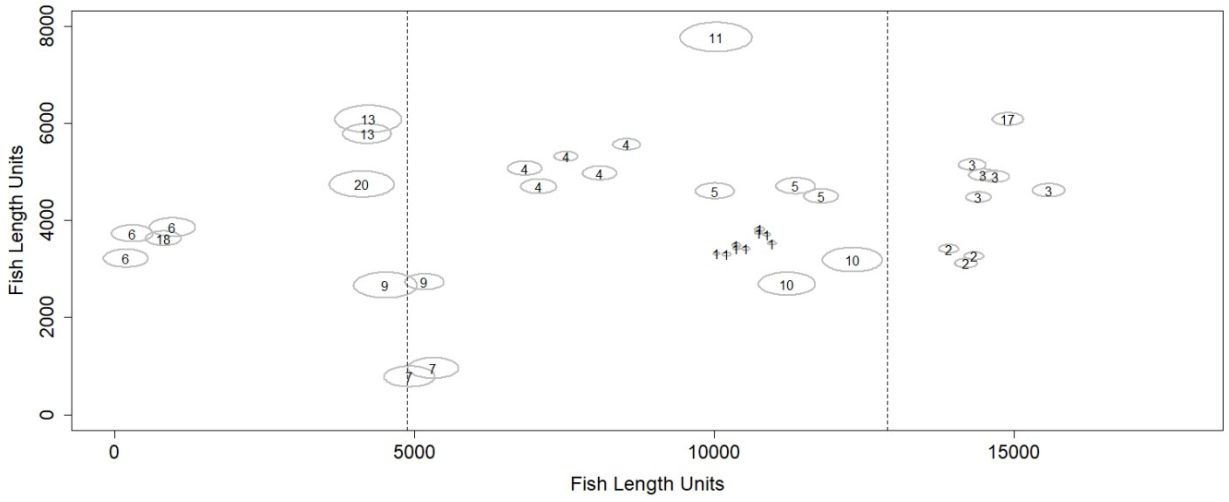
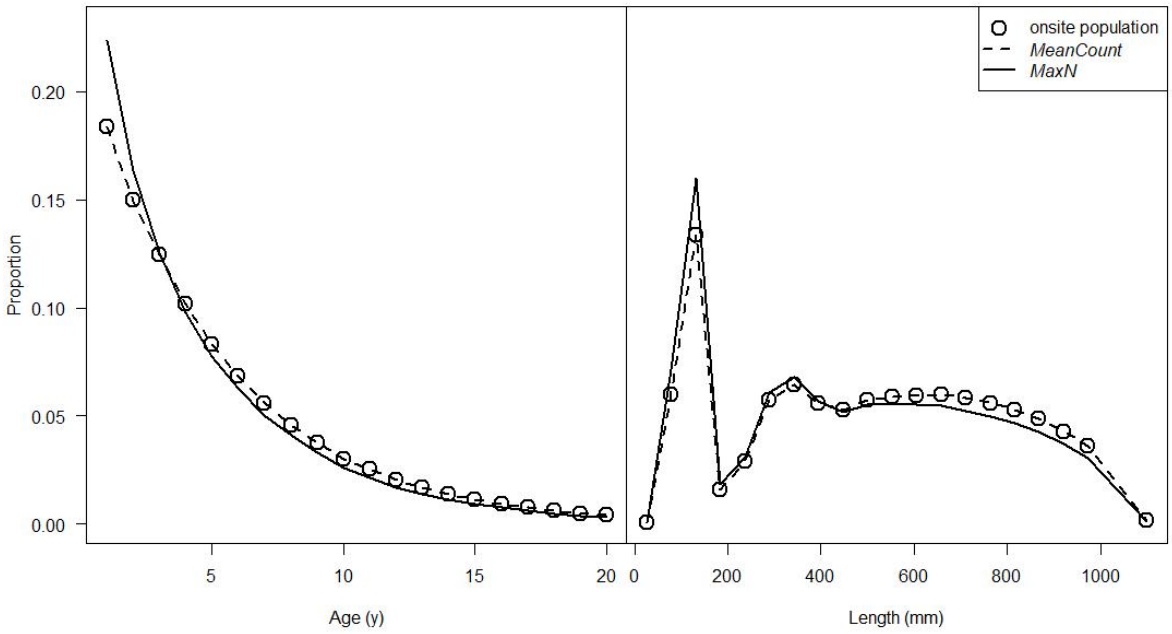
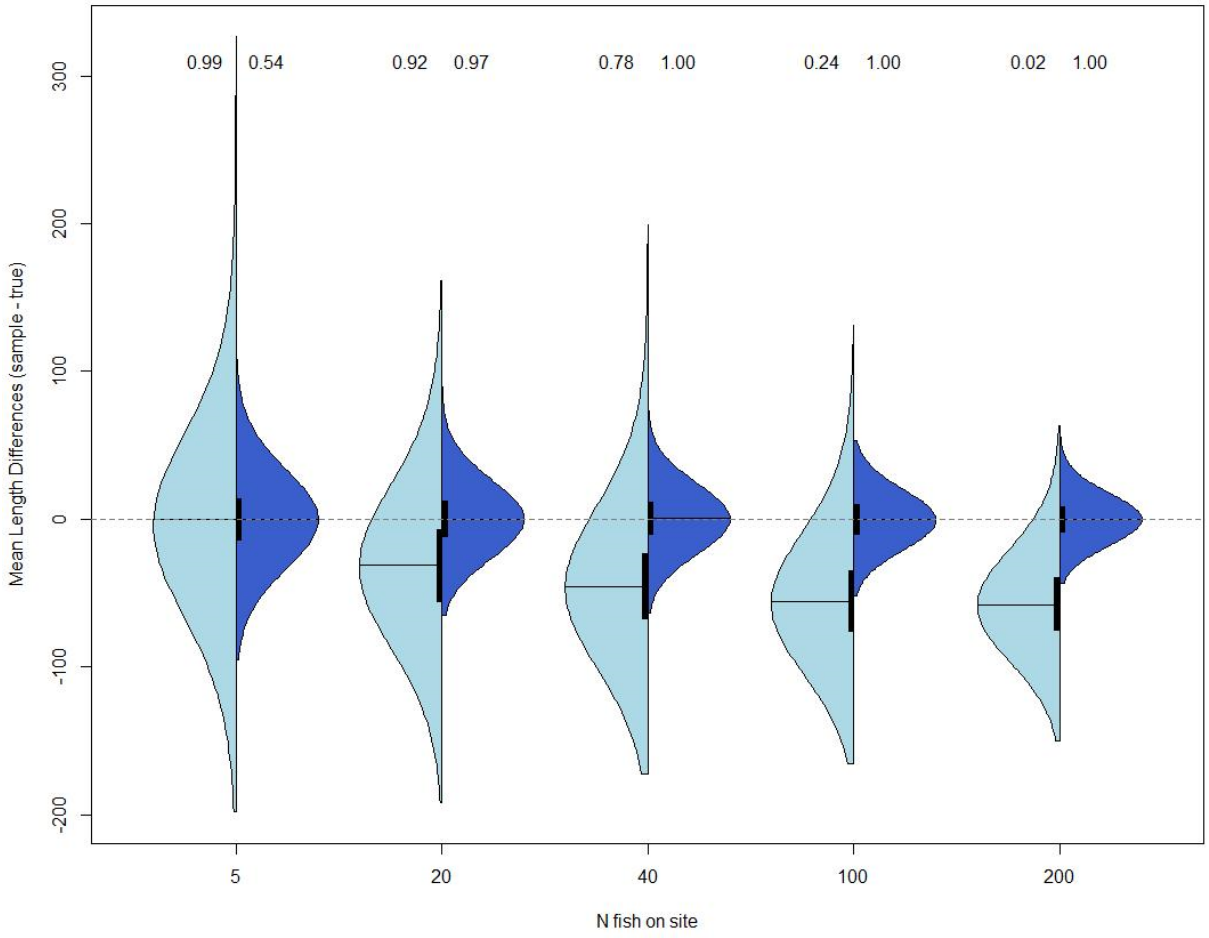



Fig. 1. An example of a single video frame from a simulated fish population on a sample site. Fish are represented by ovals with the width and height representing the relative size at age, indicated by the number in the center of the oval. The video viewing frame is delineated between the vertical dashed lines with fish outside those lines avoiding detection. Simulation parameters are set to $S_d = 1$, $N = 40$, and growth and $M$ values are set to the base values indicated in Tables 1 and 2.



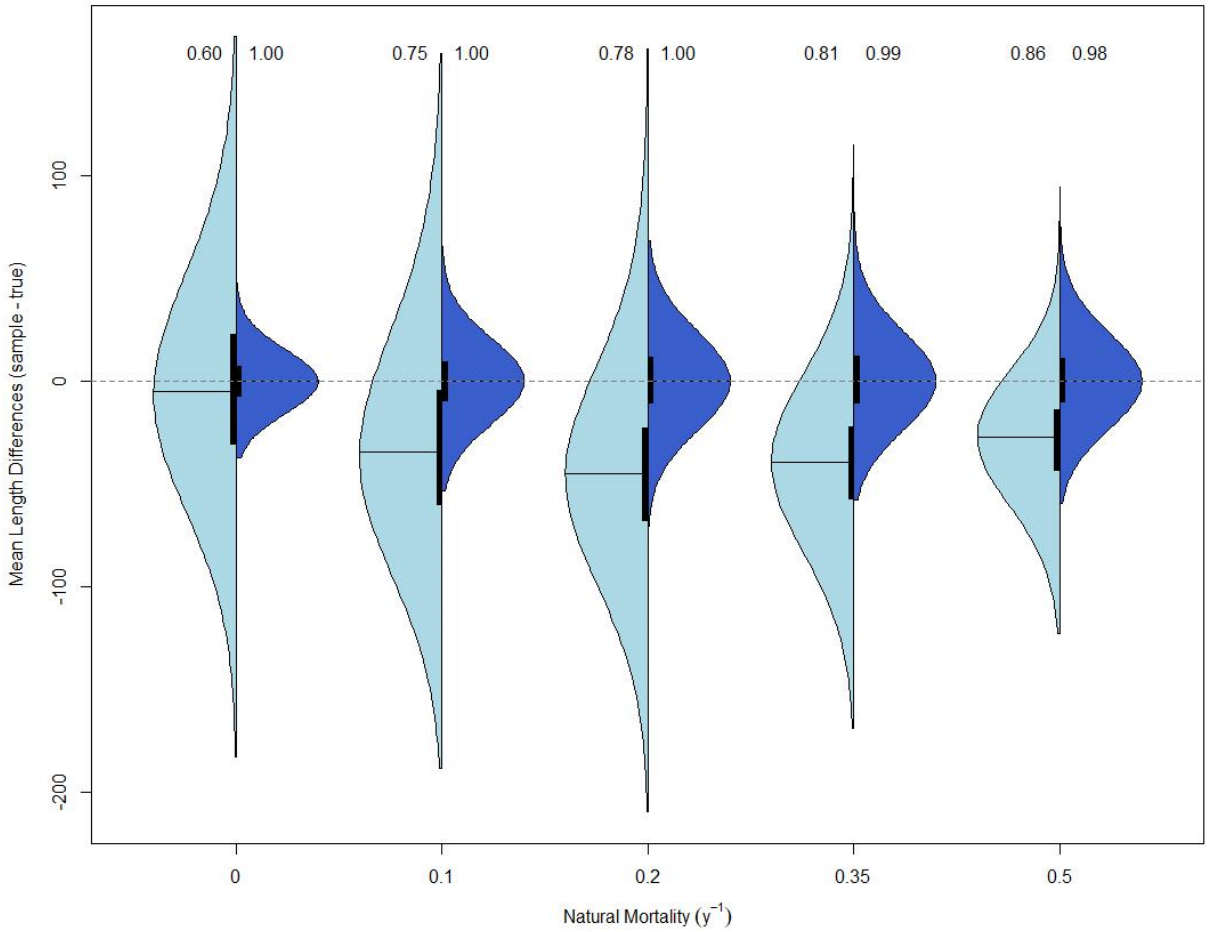

503

504 Fig. 2. Average age and length composition estimates from the video sampling simulation with
505 $N = 40$, $M = 0.2$, $S_d = 1$, $\rho = 0$, and $C_p = 0.45$ parameter settings and sampling repeated 20,000
506 times. *MeanCount* and *MaxN* age composition estimates are shown relative to the true
507 population.




508

509 Fig. 3. Split violin plots showing the distribution of differences in mean length from the sample
510 estimated and true on-site population for *MaxN* (light blue on left hand side) and *MeanCount*
511 (dark blue on right hand side) methods as a function of number of fish on site for 20,000 random
512 samples with $N = 5$ to 200, $M = 0.2$, $S_d = 1$, $\rho = 0$, and $C_p = 0.45$ parameter settings. Horizontal
513 solid black lines represent medians and the vertical bold lines represent the interquartile range.
514 Proportions along the top of each violin plot correspond to the proportion of multinomial
515 samples with similarity P-values >= 0.95 when compared to the true multinomial values.




516

517 Fig. 4. Split violin plot showing the distribution of differences in mean length from the sample
518 estimated and true on-site population for *MaxN* (light blue on left hand side) and *MeanCount*
519 (dark blue on right hand side) methods as a function of natural mortality (*M*) values for 20,000
520 random samples with $N = 40$, $M = 0.0$ to 0.5, $S_d = 1$, $\rho = 0$, and $C_p = 0.45$ parameter settings.
521 Horizontal solid black lines represent medians and the vertical bold lines represent the
522 interquartile range. Proportions along the top of each violin plot correspond to the proportion of
523 multinomial samples with similarity P-values >= 0.95 when compared to the true multinomial
524 values.

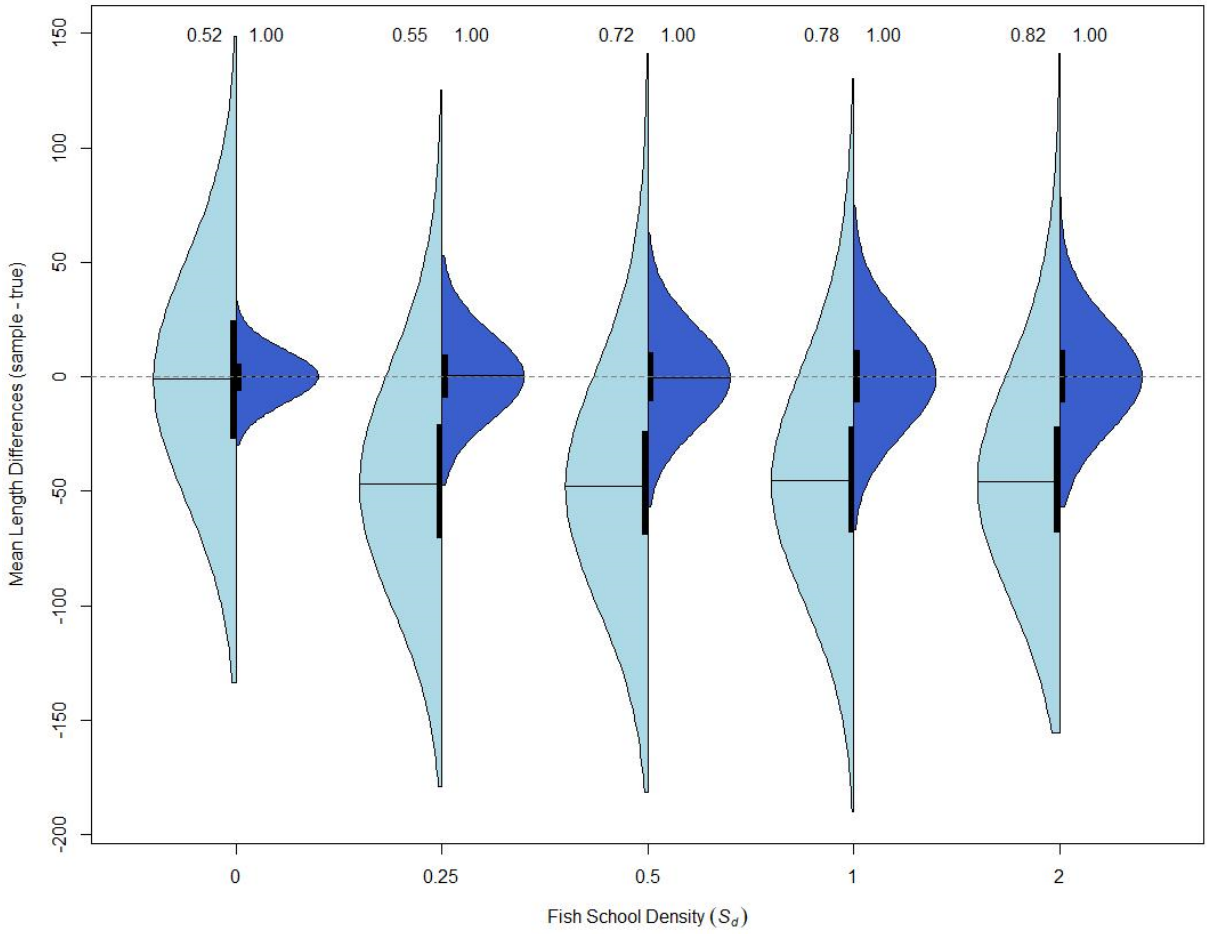
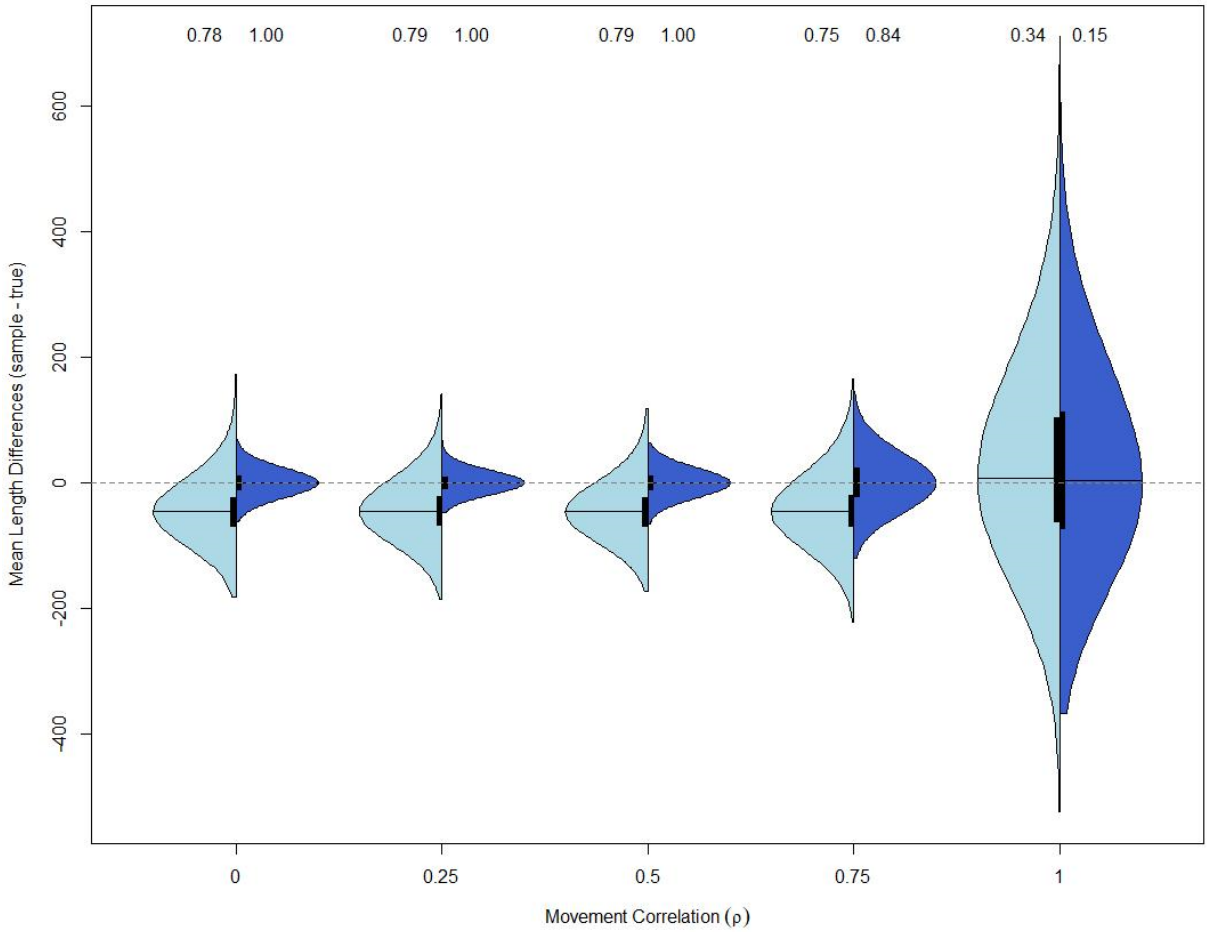
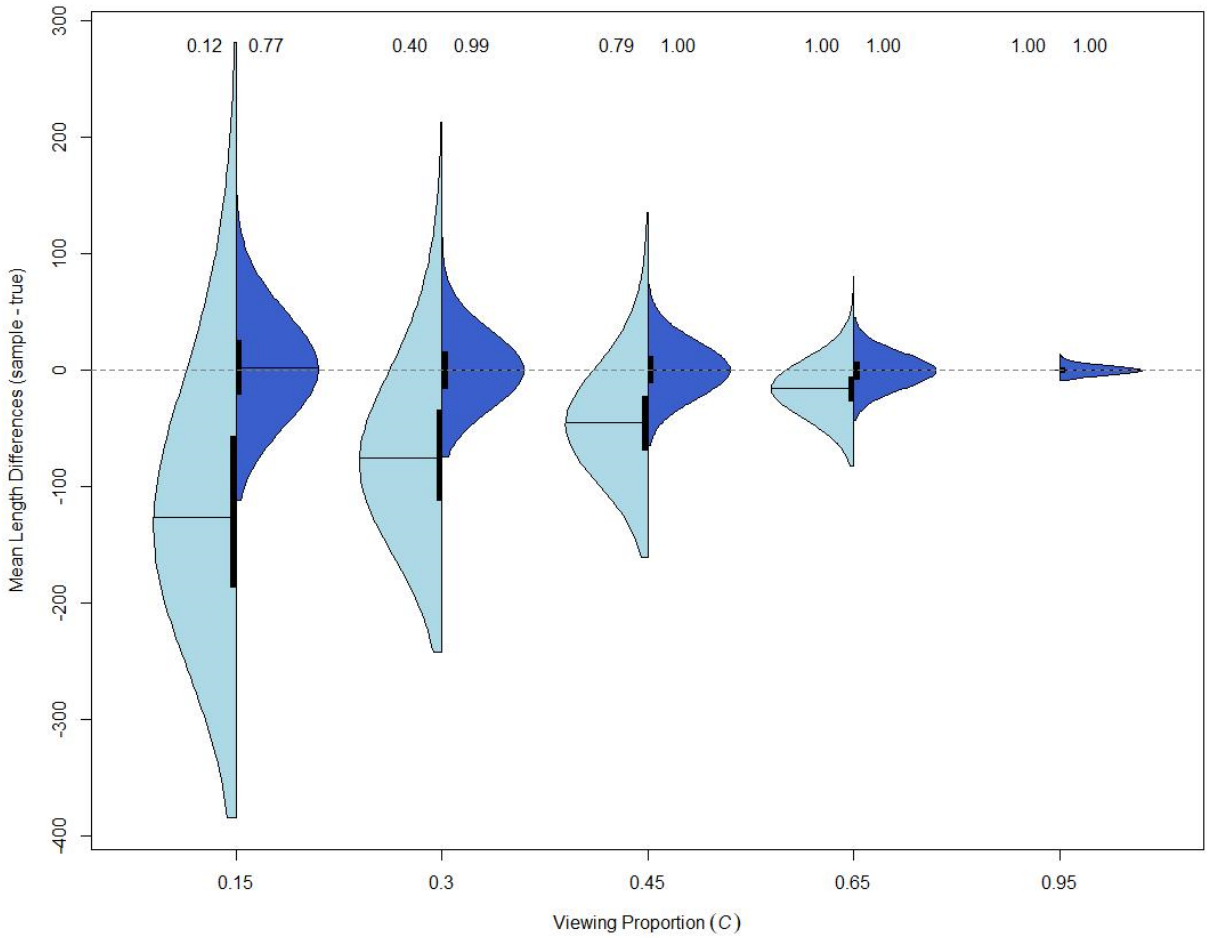



Fig. 5. Split violin plot showing the distribution of differences in mean length from the sample estimated and true on-site population for *MaxN* (light blue on left hand side) and *MeanCount* (dark blue on right hand side) methods as a function of fish school density ($S_d$) values for 20,000 random samples with $N = 40$, $M = 0.2$, $S_d = 0$ to 2, $\rho = 0$, and $C_p = 0.45$ parameter settings. Horizontal solid black lines represent medians and the vertical bold lines represent the interquartile range. Proportions along the top of each violin plot correspond to the proportion of multinomial samples with similarity P-values >= 0.95 when compared to the true multinomial values.



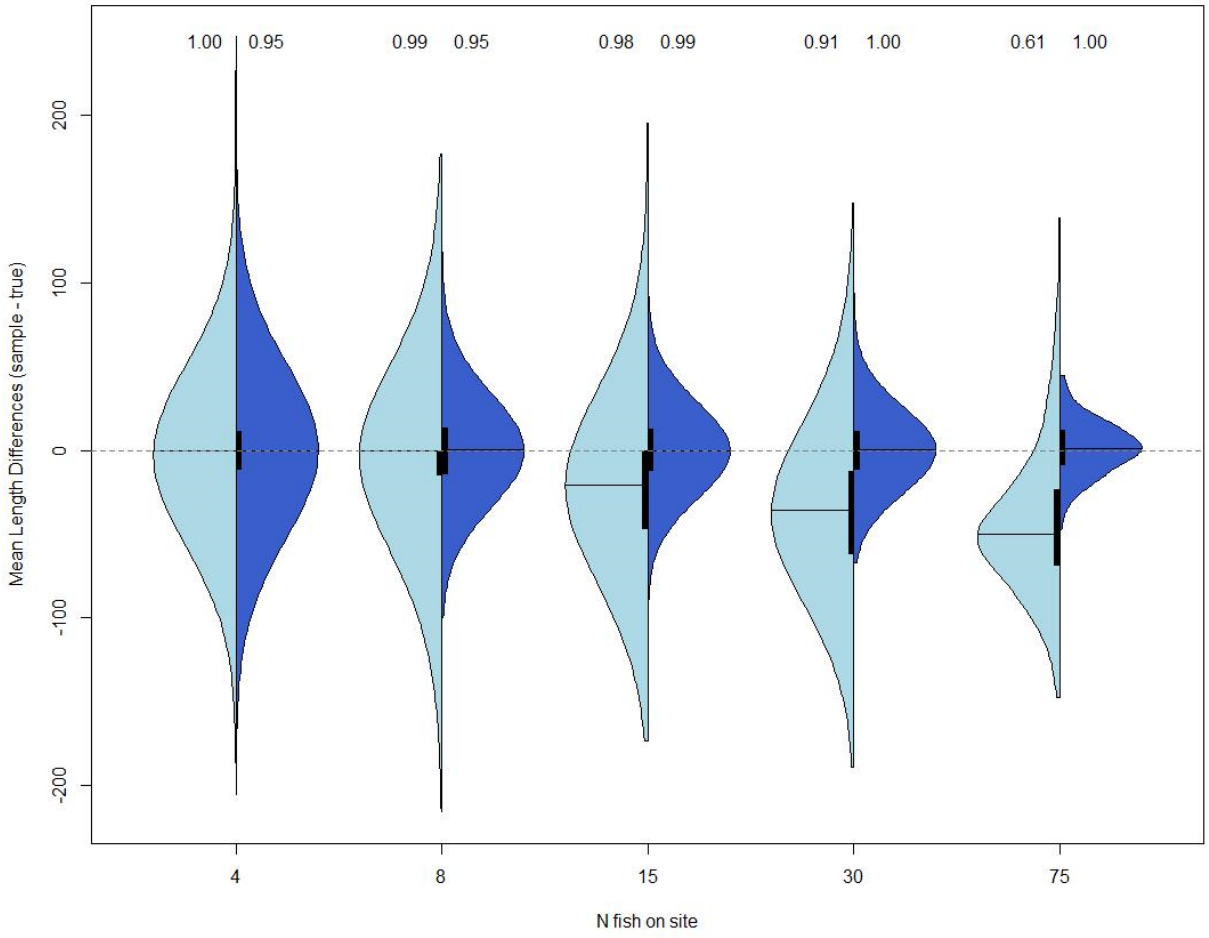

534

535 Fig. 6. Split violin plot showing the distribution of differences in mean length from the sample
536 estimated and true on-site population for *MaxN* (light blue on left hand side) and *MeanCount*
537 (dark blue on right hand side) methods as a function of movement correlation ($\rho$) values for
538 20,000 random samples with $N = 40$, $M = 0.2$, $S_d = 1$, $\rho = 0.0$ to 1.0, and $C_p = 0.45$ parameter
539 settings. Horizontal solid black lines represent medians and the vertical bold lines represent the
540 interquartile range. Proportions along the top of each violin plot correspond to the proportion of
541 multinomial samples with similarity P-values >= 0.95 when compared to the true multinomial
542 values.



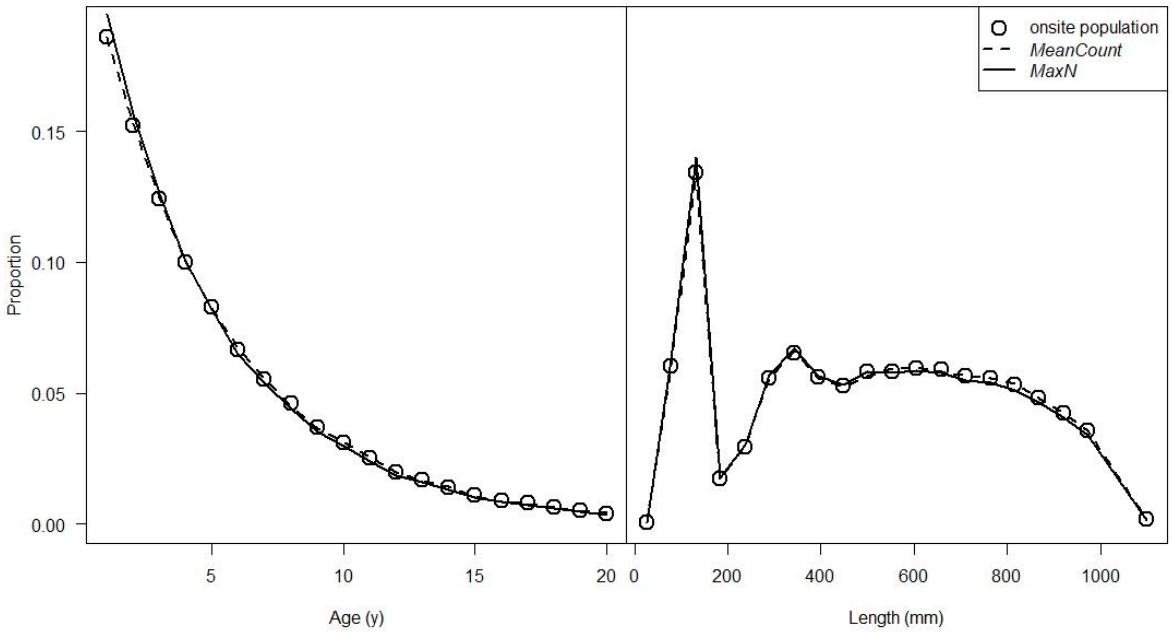

543

544 Fig. 7. Split violin plot showing the distribution of differences in mean length from the sample
545 estimated and true on-site population for *MaxN* (light blue on left hand side) and *MeanCount*
546 (dark blue on right hand side) methods as a function of video viewing proportion values for
547 20,000 random samples with $N = 40$, $M = 0.2$, $S_d = 1$, $\rho = 0$, and $C_p = 0.15$ to 0.95 parameter
548 settings. Horizontal solid black lines represent medians and the vertical bold lines represent the
549 interquartile range. Proportions along the top of each violin plot correspond to the proportion of
550 multinomial samples with similarity P-values >= 0.95 when compared to the true multinomial
551 values.




552

553 Fig. 8. Split violin plot showing the distribution of differences in mean length from the
554 simulated Red Snapper population for *MaxN* (light blue on left hand side) and *MeanCount* (dark
555 blue on right hand side) methods as a function of number of fish on-site for 20,000 random
556 samples. Horizontal solid black lines represent medians and the vertical bold lines represent the
557 interquartile range. Proportions along the top of each violin plot correspond to the proportion of
558 multinomial samples with similarity P-values >= 0.95 when compared to the true multinomial
559 values.




560

561 Fig. 9. Age and length composition estimates from the video sampling simulation of Red
562 Snapper with stochastic values matching those observed in video samples. Sampling repeated
563 20,000 times. Average *MeanCount* and *MaxN* age and length composition estimates are shown
564 relative to the true population.