



# AN EVALUATION AND POWER ANALYSIS OF FISHERY INDEPENDENT REEF FISH SAMPLING IN THE GULF OF MEXICO AND U.S. SOUTH ATLANTIC



by Paul B. Conn

U.S. DEPARTMENT OF COMMERCE  
National Oceanic and Atmospheric Administration  
National Marine Fisheries Service  
Southeast Fisheries Science Center  
101 Pivers Island Road  
Beaufort, NC 28516 USA

February 2011





NOAA Technical Memorandum NMFS-SEFSC-610

AN EVALUATION AND POWER ANALYSIS OF FISHERY INDEPENDENT REEF  
FISH SAMPLING IN THE GULF OF MEXICO AND U.S. SOUTH ATLANTIC

BY

PAUL B. CONN  
National Marine Fisheries Service  
Southeast Fisheries Science Center  
101 Pivers Island Road  
Beaufort, NC 28516 USA

U.S. DEPARTMENT OF COMMERCE  
Gary F. Locke, Secretary

National Oceanic and Atmospheric Administration  
Dr. Jane Lubchenco, Under Secretary of Oceans and Atmosphere

National Marine Fisheries Service  
Eric Schwaab, Assistant Administrator for Fisheries

February, 2011

This Technical Memorandum series is used for documentation and timely communication of preliminary results, interim reports, or similar special-purpose information. Although the memoranda are not subject to complete formal review, editorial control, or detailed editing, they are expected to reflect sound professional work.

## NOTICE

The National Marine Fisheries Service (NMFS) does not approve, recommend or endorse any proprietary product or material mentioned in this publication. No reference shall be made to NMFS or to this publication furnished by NMFS, in any advertising or sales promotion which would imply that NMFS approves, recommends, or endorses any proprietary product or proprietary material mentioned herein which has as its purpose any intent to cause directly or indirectly the advertised product to be used or purchased because of this NMFS publication.

This report should be cited as follows:

Conn, Paul B. 2011. An evaluation and power analysis of fishery independent reef fish sampling in the Gulf of Mexico and U.S. south Atlantic. NOAA Technical Memorandum NMFS-SEFSC-610, 38 p.

This report will be posted on the SEFSC website later in 2011 at  
URL: [http:// www.sefsc.noaa.gov/](http://www.sefsc.noaa.gov/)

Copies may be obtained from:

Paul Conn  
National Marine Fisheries Service  
NOAA Alaska Fisheries Science Center  
National Marine Mammal Laboratory  
7600 Sand Point Way NE, Building 4  
Seattle, WA 98115-6349  
Voice: 206-526-4235  
FAX: 206-526-6615  
Paul.Conn@noaa.gov

Also available for purchase in paper copy and microfiche form  
National Technical Information Service (NTIS)  
5285 Port Royal Road  
Springfield, VA 22161  
1-800-553-NTIS  
<http://www.ntis.gov>

## Executive summary

I investigated the power of reef fish surveys in the Gulf of Mexico (GOM) and U.S. south Atlantic (SA) to detect population changes for several species of commercial interest. Related to this goal, I also investigated the performance of different underwater camera metrics in indexing abundance, and examined how the annual coefficient of variation (CV) of model-based relative abundance indices changed as a function of sample size in the SA. Major findings included the following:

- The commonly used “MinCount” statistic for underwater video data (the maximum number of individuals of the target species observed on any one frame of video) may provide biased estimates of population trend, as this summary does not scale linearly with true abundance. In contrast, the mean count of focal taxa across all video frames was found to be unbiased and resulted in similar CV to the “MinCount” statistic. Researchers should seriously consider using this statistic in the Gulf and elsewhere for indexing abundance. However, care may be needed to avoid sampling during ‘transient’ periods where fish are moving into and out of the immediate proximity of the trap.
- In GOM, the power to detect population trends for gag, red snapper, and red grouper were poor for three year time series, low for five-year time series and moderate-high for ten-year time series. In general, the ability to detect population changes was more dependent on the length of the time series than on the range of sample sizes considered. Lack of complete spatial coverage did not appreciably affect power or lead to biased inferences, but I caution that this is likely due to the way in which simulations were structured. In particular, simulations did not permit the types of complex space-time interactions often observed in real world data sets.
- In SA, the power to detect population trends for black sea bass and red porgy was quite high, where coefficient of variation (CV) of annual model based indices were predicted to be under 0.2 with current MARMAP sample sizes.

- In SA, the power to detect population trends for red snapper from fishery independent (MARMAP) data is extremely low, due to the relative infrequency with which they are observed, and the ‘messiness’ of the data (e.g., there is substantial zero inflation and overdispersion present in the MARMAP data). Replicating these factors in simulations, median annual model-based CVs were predicted to be above 2.0 for 5-year time series, even if some effort was shifted to north Florida and current MARMAP sample sizes were multiplied by 10. Although the situation may improve if the proportion of positive MARMAP sets increases as the red snapper stock recovers, managers should prepare for the situation where fishery independent data are quite noisy, with annual data points that are relatively meaningless if taken individually. Only by examining a relatively long time series will it be possible to discern increases or decreases in the red snapper stock.
- Future efforts should focus on appropriate levels of subsampling in fishery independent surveys. For SA red snapper, there was evidence of considerable overdispersion, which indicates structural deficiencies in the models fit to MARMAP data. Part of this deficiency likely lies with the way in which MARMAP data are collected - for instance, by collecting a large number of samples within a relatively small spatial range (i.e., subsampling). The tradeoff between collecting subsamples (which increases the number of samples that can feasibly be collected) and collecting samples that have a greater degree of spatial independence deserves further investigation. Ideally, this investigation would occur using spatially explicit statistical modeling to indicate likely levels of spatial autocorrelation.

# 1 Introduction

Marine fishery managers have been incorporating time and spatial closures into their repertoire of regulation alternatives in the south Atlantic and Gulf of Mexico. Although such management can prove effective, it poses some serious challenges for the assessment of marine resources, as it decreases the utility and breadth of fishery dependent sampling. Since stock assessments crucially depend on estimates of relative abundance and age-structure from the entire range of a stock, this suggests that stock assessments will need to rely much more heavily on data gathered from fishery independent surveys.

One issue with current fishery independent sampling programs is that standardized catch per unit effort (CPUE) time series are often noisy, hampering their utility. For example, a 2008 internal review of National Marine Fisheries Service (NMFS) Southeast Fisheries Science Center (SEFSC) fishery independent surveys conducted in the Gulf of Mexico concluded that sampling effort would need to be quadrupled in order to achieve CVs on the order of 20% (the current level varies by species and survey but CVs are typically in the 30-50+% range). The reviewers also recommended increasing the spatial extent of their current sampling frame for a number of surveys (including the SEAMAP reef fish video/trap survey). The situation is worse in the south Atlantic, with current MARMAP surveys resulting in CVs that are often too high to be practically useful. For instance, MARMAP survey data were not recommended for use in either the SouthEast Data, Assessment, and Review (SEDAR) 15 assessment or SEDAR 24 assessments of red snapper because the number of samples was deemed inadequate for indexing relative abundance.

For these reasons, the SEFSC is working to increase the scope of fishery independent sampling in the south Atlantic, and possibly the Gulf as well. Since resources are limited, a crucial question is where and how to increase sampling efforts. Although the recent internal review of fishery independent sampling in the Gulf of Mexico provides some intuition about desired levels of precision in relation to sampling effort, more guidance would be provided by a formal power analysis that explores the probability of detecting a ‘significant’ trend in abundance over set time frames given different levels of sampling effort. This exercise could also inform survey designers on possible tradeoffs between expanded spatial coverage and sample sizes at existing locations.

As reef fish are the primary target for spatial closures, I focus my analysis on reef fish surveys as they are currently conducted in the Gulf of Mexico and U.S. south Atlantic. In particular, I focus on the Gulf SEAMAP video survey and south Atlantic MARMAP

Chevron trap survey. However, since current proposals for expanding surveys in the south Atlantic include incorporation of video sampling methods, it may be useful to first review existing methods for analyzing video data. I thus start my analysis by examining the performance of different camera monitoring metrics. I then conduct a power analysis which looks at the power to detect population change as a function of sampling effort, spatial coverage, and underlying heterogeneity in fishery dynamics. I conduct analyses for three recreationally and commercially important species for which data are available from the Gulf of Mexico - gag grouper (*Mycteroperca microlepis*), red grouper (*Epinephelus morio*), and red snapper (*Lutjanus campechanus*). In the south Atlantic, I examine data from black sea bass (*Centropristis striata*), red porgy (*Pagrus pagrus*), and red snapper. In addition to a formal power analysis, I also compute median CVs associated with model-based estimates of annual relative abundance for the south Atlantic for the three stocks of interest.

## 2 Analyzing camera data

Current SEAMAP reef fish video surveys in the Gulf of Mexico utilize two stage sampling, in which large grid cells (Fig. 1) containing reef fish habitat (10 minute by 10 minute blocks) are sampled using a stratified random design (stratification is based on region and by reef habitat area). Secondary sites within blocks are then selected randomly. For each sampled site, a baited array of four, orthogonally arranged video cameras is dropped to the ocean floor (current plans are to switch to stereo cameras which also allow fish to be measured). After the camera settles, the sampling duration is 20 minutes. Once researchers return to the lab, data are analyzed by randomly selecting a single camera from each sample that has focused, unobstructed views. Three quantities are then computed for each site sampled: 1) presence/absence of focal species, 2) Maximum Count (MaxCount; each fish is counted every time it enters the video frame), and 3) Minimum Count (MinCount; the greatest number of individuals of the species seen in a single frame). Since each individual can be counted more than once using the MaxCount approach (for instance, when it leaves and then reenters the field of view), MaxCount may actually exceed the number of individuals present at a given sample site. However, MinCount almost assuredly misses a large number of individuals that are out of view of the camera. Recent stock assessments in the Gulf of Mexico (e.g., SEDAR 7, red snapper) have utilized MinCount when constructing indices of abundance (for instance, using GLMs).

Further inspection of video summary statistics suggests that the true relationship



between abundance and presence/absence is nonlinear. For instance, Royle and Nichols (2003) noted that the relationship between the probability of detecting a species at a site  $i$  ( $p_i$ ) was related explicitly to both abundance at the site ( $N_i$ ) and individual detection probability  $r$  through the equation

$$p_i = 1 - (1 - r)^{N_i} \tag{1}$$

Assuming a constant probability of detection for individuals ( $r$ ), the relationship between observed occupancy and abundance is thus a nonlinear function of abundance (Fig. 2). Given that relative abundance is often modeled as a linear function of true abundance within stock assessment models (e.g.,  $CPUE_t = qN_t$ ), use of proportion of sites where the focal species is detected is clearly a poor choice. Alternatively, hierarchical modeling using the above relationship in Eq. 1 might be used to construct an index using presence/absence data. An advantage of this approach is that one can directly assess the possibility that catchability changes over time. However, a subset of sites would need to be visited more than once each year in order for parameters to be identifiable.

The MinCount statistic was first proposed by Ellis and DeMartini (1995) for indexing reef fish, and shown to outperform several other possible metrics. Willis et al. (2000) compared MinCount video statistics to numbers counted in visual surveys and in angling surveys, and found that the three were correlated. However, there is little published information comparing coefficient of variation (CV) for alternative video metrics.

I conducted a small simulation study to compare bias and CV for several monitoring metrics, including MinCount, MaxCount, ‘‘Snapshot’’ which utilized one frame from the end of each video time series, and ‘‘MeanCount’’ which was calculated by computing the average number of fish present in the video frame over the course of a simulated time series. To my knowledge the latter two metrics have never been evaluated before. I simulated data for a hypothetical sampling site as a function of  $N_i$  and some assumed parameters describing the underwater observation process. Specifically, I assumed that the proportion of fish present at site  $i$  that are initially in the field of view was binomially distributed with index  $N_i$  and success probability  $\pi_i$  and that the movement of fish in and out of the field of view could be described by a Poisson process with exponentially distributed waiting times. The per minute rate at which fish that are ‘‘off camera’’ come into the field of vision is given by  $\alpha_1$ , while the per minute rate at which fish that are in the field of vision leave the field of vision is given

by  $\alpha_2$ . Two scenarios were considered in simulations: an  $\alpha_1 = 0.05, \alpha_2 = 0.2$ , simulation, which corresponded to a fish with relatively slow movements (e.g., a grouper) and a camera with a relatively small field of vision (hence the movement rate onto the camera’s field of vision is smaller than the movement rate out of the field of vision), and an  $\alpha_1 = 0.5, \alpha_2 = 0.5$  simulation which tried to mimic a fish with faster movement rates and a camera with a wider field of vision. Stochastic simulations of this assumed process (cf., Renshaw, 1991) suggested that all metrics except for MinCount were unbiased (Fig. 3). The MinCount statistic underestimated the the population trend; for a twenty-fold increase in abundance, MinCount underestimated the true population change by 49% for the  $\alpha_1 = 0.05, \alpha_2 = 0.2$  simulation and by 25% for the  $\alpha_1 = 0.5, \alpha_2 = 0.5$  simulation.

On the other hand, the MinCount metric typically had the smallest CV (as calculated separately for each design point) (Fig. 4). The MeanCount metric also performed well with respect to CV in both simulations, while precision of the MaxCount statistic could be good if the rates of fish moving in and off camera are reasonably low (as in Fig. 4A) or poor for cases where fish movement rates are high (e.g., Fig. 4B).

If one assumes that all fish behave independently and move in and off the camera’s field of vision at constant rates (as was assumed in simulations), it is easy to show that relative abundance trends from the MeanCount and Snapshot approaches are unbiased. However, simulation results have suggested that MinCount tends to underestimate the degree of change in abundance values. Although difficult to examine analytically, this is not surprising as MinCount is the maximum order statistic for the stochastic process, and it is not intuitive that this quantity would scale linearly with true underlying abundance. As a motivating example, consider two frames from a single video that are separated by enough time that observations can be assumed to be independent. In this case, the observed counts are given by two binomial draws. Now consider two different sites, one of which has  $N = 1$  fish present, and another that has  $N = 2$  fish present. The expected MinCount statistic for the  $N = 1$  site is 0.75, which represents 75% of true abundance. In the  $N = 2$  site, the expected MinCount is 1.3125, which is 65.6% of true abundance. Had we used MinCount to infer the differences in relative abundance between the two sites, we would expect to come to the erroneous conclusion that the  $N = 2$  site had 75% more fish than the  $N = 1$  site.

Although we have found reason to prefer the MeanCount procedure on both theoretical and applied grounds, we do note that the biases reported here are likely larger

then would be obtained if delta-generalized linear model (delta-GLM) standardization procedures (cf., Lo et al., 1992) were used to standardize MinCount summaries for real fish populations. This is because delta-GLMs (cf., Lo et al., 1992) analyze both the delta (presence/absence) component, as well as abundance at a site given presence. In this exercise, we have only attempted to examine bias in the latter component. Still, the simulation exercise does demonstrate potential for bias when estimating trends from MinCount statistics; the degree of bias will probably be most pronounced in species that are highly prevalent (that is, have a lower proportion of zero counts).

## 3 Simulation study

### 3.1 Gulf of Mexico

To assess the relationship between sample size and power to detect population trends, I first gathered data from the SEAMAP reef fish survey for the Gulf of Mexico, and MARMAP Chevron trap data for the U.S. south Atlantic. The SEAMAP data were summarized by year, strata, and soak time, and reported the video metric “Min-Count.” Original video tape viewing procedures used a time in - time out (TITO) protocol which would have been sufficient to compute other metrics (e.g., MaxCount, MeanCount); however, this protocol was switched in 2008 to one in which only Min-Count could be computed. The SEAMAP data gave the number of the focal species that were observed in individual videos by year. I focused effort on three species of prominence in the Gulf: gag grouper, red grouper, and red snapper.

Simulations were designed to answer a number of questions pertinent to past and future fishery independent sampling efforts. First, is there a trade off between expanding sampling coverage versus intensity of samples in currently sampled locations? The answer to this question likely depends on whether or not there is spatial heterogeneity in population trend. Second, what levels of sampling effort would need to be exerted to detect population changes? For instance, what levels of sampling effort would be needed to have 80% power to detect a population increase of 10% over five years?

In conducting a power analysis, it was desirable to retain the “messiness” of typical survey data, while still being able to impose desired population trends. Two common problems with survey count data are (i) overdispersion (higher than expected variance relative to standard parametric statistical distributions), and (ii) zero-inflation (an over-abundance of zeros relative to standard parametric distributions). To ad-

dress these two factors, a common approach in fisheries index standardization is to use delta-GLMs (cf., Lo et al., 1992). Here, I used a similar approach based on zero-inflated, generalized Poisson (ZIGP) count models (Czado et al., 2007) to capture the “messiness” of typical data sets. These models are parameterized as a finite mixture distribution, where the counts at an unknown proportion  $(1 - \omega)$  of the sampled sites are assumed to follow an overdispersed Poisson process, while counts at the remaining proportion  $(\omega)$  of sites are assumed to be zero. I started by fitting this model to gag, red grouper, and red snapper datasets using the package ZIGP in the R Programming Environment (R Development Core Team, 2007). In particular, I used MinCount as the response variable and assumed that the log-transformed Poisson mean parameter  $(\log(\mu))$  was dependent on strata, while the zero inflation parameter  $(\text{logit}(\omega))$  and overdispersion  $(\log(\phi))$  parameter were constant over space and time. Standard regression diagnostics suggested the model fit the data reasonably well. Resulting parameter estimates are given in Table 1.

Using this approach, the expected count at a given site is given by  $(1 - \omega_t)\mu_{st}$ , where the subscript  $t$  indexes time (e.g., year), and the subscript  $s$  indexes strata. This expectation provides some intuition on the considerations needed to model population increases. First, it is apparent that increases can result from either a decrease in  $\omega$  (the proportion of extra zeroes), or increases in  $\mu$ . Implementing either of these scenarios is relatively straightforward in the case that the population in each strata is increasing at the same rate. However, if there is spatial heterogeneity in abundance trends, additional elaborations are needed.

Simulations were conducted using a factorial design, with the following factors varied:

- spatial coverage (50% or 100%)
- spatial heterogeneity in abundance trends (yes/no)
- sample sizes (half, current, or double of “average” levels [assumed to be 420 sets/year])

To investigate the effect of incomplete spatial coverage, I fit a normal distribution to strata-specific estimates of process intensity  $(\mu_s)$  from fits to the SEAMAP data, and used the resulting distributions (Fig. 5) to sample strata effects for additional strata. All seven currently defined spatial strata were assumed to be sampled in all simulation scenarios; an additional seven were assumed to exist (but were not sampled) in

simulations where spatial coverage was assumed to be 50%.

Changes in abundance were implemented by manipulating  $\mu_{st}$  while keeping the zero-inflation parameter constant at its estimated value for each species. All simulations used the same general structure:

1. Starting with the initial Poisson intensity  $\mu_{s1} = \mu_s$  estimated from SEAMAP data, simulate time series of strata-specific Poisson intensities as  $\mu_{s,t+1} = \mu_{st}\lambda_s \exp(Z_t)$ , where  $\lambda_s$  gives a deterministic rate of population increase, and  $\exp(Z_t)$  gives log-normal process error. I assumed that  $Z_t \sim \text{Normal}(0, 0.01)$  for all simulations.
2. Simulate data for sampled site  $i$  in strata  $s$  and time  $t$  as  $X_{ist} \sim \text{ZIGP}(\mu_{st}, \phi, \omega)$ , where  $\phi$  and  $\omega$  are estimated from the SEAMAP data.
3. Fit a ZIGP model with constant overdispersion and zero-inflation to the simulated data, assuming the model  $\log(\mu_{st}) = \beta_s + \gamma t$ . A Bernoulli response variable was recorded for whether or not a 90% asymptotic confidence interval for  $\gamma$  included zero; proportional population change was also recorded, and was calculated as  $\Delta = \sum_s \mu_{sT} / \sum_s \mu_{s1}$ .

Variation in abundance trends between simulation replicates was obtained by drawing  $\lambda$  from a  $\text{Normal}(1.0, 0.01)$  distribution for each simulation. Simulations without spatial heterogeneity in abundance used the relation  $\lambda_s = \lambda$ , while those including spatial heterogeneity included spatial variation in the form  $\lambda_s \sim \text{Normal}(\lambda, 0.01)$ . For each design point, I simulated 1000 ten year time series.

Results for all three species (Figs. 6-8) indicated that the power to detect population trends was virtually nonexistent for three year time series, low for five year time series, and moderate-high for ten year time series. The length of the time series generally had a much larger influence on power than did sample sizes; the power to detect trends was highest for red snapper, followed by red grouper and then gag. Interestingly, results were largely consistent between simulation scenarios - i.e., it did not seem to matter whether sample coverage was 50% or 100% or whether there were different relative abundance trends in different spatial strata. In all cases, type I error rates ( $\alpha$ ) were close to nominal. However, since the deviation of each strata from the mean trend was drawn from a normal distribution, it is worth noting that the strata that were sampled were largely “representative” of strata that were not sampled. I expect that results would have been substantially different had sampled strata exhibited a different mean trend than unsampled strata. In particular, I would expect realized  $\alpha$

levels to be higher than nominal. Thus I caution against using these results to justify limiting the spatial scope of the survey.

### **3.2 South Atlantic (U.S.)**

For the U.S. south Atlantic, I conducted a similar analysis to the Gulf, but used MARMAP Chevron trap data for black sea bass, red porgy and red snapper. The first two species are generally recognized as being amenable to trapping; in contrast, red snapper are caught much less frequently in traps (this is likely due to low relative abundance in sampled sites; there is ancillary evidence from the Gulf of Mexico that they do trap well). The MARMAP data is currently the largest and longest running fishery independent collection program for reef fish in the U.S. south Atlantic. However, data are often too sparse to provide adequate indications of relative abundance trends. As such, scientists and managers are interested in expanding the survey. Possible expansions include (i) increasing the geographic range (particularly to north and south), (ii) increasing sample sizes and the overall sampling universe in currently sampled areas (particularly in north Florida hard bottom habitats where red snapper landings are the highest), and (iii) expanding the gears used to include camera/video traps to increase catchability for ‘hard to trap’ species such as gag grouper. However, traps will likely remain a feature of most surveys to facilitate collection of biological samples.

To investigate possible effects of range and sample size expansions on the ability to detect population trends from MARMAP Chevron trap data, I gathered existing Chevron trap data from 1988-2009 for the three focal species. Initial fits to the data using a GAM with quasipoisson error structure (Fig. 9-11) indicated that catch rates varied by a number of factors, including year, latitude, depth, month, and soak duration. Sample sizes varied substantially between species; out of a total of 7,327 sets, 155 caught red snapper (a total of 388 fish were caught), while 2,793 (total 15,229) caught red porgy, and 2,908 (total 78,412) caught black sea bass. Although it is difficult to read too much into trends in the red snapper data because of low sample sizes, it is apparent that there are commonalities and differences between the three species. For instance, the predicted catch for each species was highest between April and August/September than other months. Hence, expansion of sampling effort into other times of the year (e.g., by using existing sampling platforms for an extended period of time) is unlikely to result in substantial increases in sample size. Also, it is apparent from red porgy and black sea bass plots that catch rates decrease with trapping duration past some threshold value; however, this may be due to a correla-

tion between trapping duration and weather events. Expanding sample coverage into deeper waters than usually covered (e.g., 50 or 75 meters) could also be expected to result in fewer proportional sets with black sea bass, although it may result in greater catch of red porgy (and might allow one to test hypotheses about differences in age structure by depth). Finally, increased latitudinal spatial coverage would likely affect each species differently. For instance, a southern expansion may result in high catch rates for red snapper and red porgy but low catch rates for black sea bass, while a northern expansion would likely favor black sea bass over the other two species.

Although investigation of differential age structure by depth may be an interesting hypothesis to test in it's own right, I focus here exclusively on the power to detect population trends using different assumptions about expansion of sampling effort. Because there are no data with which to gauge the impact of using camera/video traps (which potentially could increase sample sizes for species that do not trap well), I focus on expansion of effort using MARMAP traps only. Current proposals for increasing sampling effort and coverage (T. Kellison, Southeast Fisheries Science Center, 101 Pivers Island Rd, Beaufort, NC 28557, personal communication) call for increasing the survey range to the north and south, and also to increase the number of samples and sample universe in hard bottom habitats in northern Florida. I thus considered the following simulation scenarios, which were implemented separately for each species:

- 1. “Status quo,” which used contemporary levels of sampling effort (the number of sets per year was set to 344, which was the three year running average for 2007-2009), drawing sample locations each year without replacement from previously sampled MARMAP sets
- 2. “Expansion 1” which simulated a range expansion of the survey so that it sampled from 25° N latitude to 37° N latitude (approximately south Florida to Maryland). This scenario assumed status quo sample sizes but drew the latitude of hypothetical sampling locations from a Uniform(25,37) distribution
- 3. “Expansion 1+” was the same as Expansion 1, but had sample sizes increased to twice the status quo level
- 4. “Expansion 2” was the same as Expansion 1+, but had 25% of it's samples diverted to increase sampling in north Florida. In this scenario, 75% of samples had their latitude drawn from a Uniform(25,37) distribution, while 25% of samples had their latitude drawn from a Uniform(27,30) distribution.

For scenarios 2-4, depth, month, and sampling duration were sampled randomly with replacement from the MARMAP database. As in Gulf simulations, a ZIGP model was

fit to existing MARMAP data, and data at proposed sample locations were generated from the fitted model by conditioning on a particular level of latitude, depth, sampling duration, and month.

For each species, I fitted a ZIGP model to the MARMAP data in which

$$\log(\mu) = \text{Month} + \text{Month}^2 + \text{Duration} + \text{Duration}^2 + \text{Latitude} + \\ + \text{Depth} + \text{Depth}^2 + \text{Year}$$

and where  $\omega$  and  $\phi$  were assumed constant. The original GAM fits (i.e., Figs 9-11) were used to suggest an appropriate level of complexity for each term, with the exception of year and latitude. For year, I used a linear trend since it's only real purpose was to generate a reasonable starting value for 2009, the year chosen for a baseline value to start simulations. As there were relatively few MARMAP samples in the latitudinal extremes (Fig. 12), I modeled latitude as a categorical variable, with the following categories: 27-30 (North Florida; NFL), 31 (Georgia; GA), 32-33 (South Carolina; SC), and 34 (North Carolina; NC). The same process of simulating and analyzing data was used as in the Gulf analysis (i.e., using steps 1-3 in the previous section). However, I did not consider any scenarios with spatial heterogeneity in underlying population trend; results from this experiment should thus be viewed as a 'best case scenario.' For scenarios that expanded sample coverage to previously unsampled latitudes, I assumed that the closest latitudinal stratum best represented abundance (i.e., relative abundance in latitudes  $> 34$  was equal to that in NC, and abundance in latitudes  $< 27$  was equal to that of NF).

Simulations revealed that power to detect population trends was reasonable for black sea bass and red porgy (Figs. 13-14), but poor for red snapper (Fig. 15). Increasing sample coverage alone did not greatly increase power; however, increases in sample size and disproportionately high sampling in northern Florida did help. Regardless, power to detect population changes for red snapper remained low. Under the most optimistic scenario (increasing coverage and doubling sample sizes), the power to detect an increase or decrease of 50% was around 40% for 3- and 5-year time series and around 60% for a 10-year time series.

In addition to power, I also calculated average annual sample sizes for red snapper caught in MARMAP traps during the first three years of simulations. For the 'Status quo' scenario, an average of 12.5 red snapper were caught per year in an average of 4.9 independent sample events (the number of distinct sample events are often used



as the 'effective' sample size when modeling age compositions because ages from those caught simultaneously are usually not statistically independent). For the 'Expansion 1' scenario, these numbers increased to 16.7 and 6.4, respectively; for the 'Expansion 1+' scenario, they were 32.3 and 12.7; for the 'Expansion II' scenario, they were 36.6 and 14.3. Thus, in the best case scenario, the effective sample size for age compositions would be around 14.3/year. This level is less than the minimum threshold of 20-40 that has been used in several SEDAR assessments for determining whether age samples are actually used in the modeling process.

Another quantity of interest for the U.S. south Atlantic is the annual CV associated with a relative abundance index value. In order to calculate this value, I simulated five year time series for the 'Expansion 2' scenario, assuming that there was no trend in relative abundance save for that caused by random annual fluctuations (i.e., I assumed  $\lambda_s = 1$ ). For black sea bass and red porgy, I calculated CV assuming annual sample sizes (number of MARMAP sets) of 100, 200, . . . , or 1000. For red snapper, I calculated CV assuming annual sample sizes of 500, 1000, . . . , or 5000. For each simulation, the relative abundance index was extracted by substituting population marginal means (Searle et al., 1980) for independent variables and computing the index in year  $t$  as  $(1 - \omega)\mu_t$  (where  $\omega$  is the zero-inflation term in the ZIGP model). Coefficient of variation associated with the index was then calculated using the delta method (Seber, 1982).

Simulations suggested that annual CVs for black sea bass and red porgy would be below 0.2 for current levels of MARMAP sampling effort in the south Atlantic, a threshold often targeted for relative abundance indices (Fig. 16). In contrast, simulations suggested that MARMAP sampling for red snapper could be expected to yield model based CVs above 2.0, even if sampling effort were multiplied by ten (Fig. 17). This was largely a result of the extreme levels of overdispersion exhibited in MARMAP catch data for red snapper (the overdispersion term in the ZIGP model was estimated at 2.5). For instance, few sets actually caught red snapper, but those that did occasionally caught large numbers. If such trends continue, it will be extremely difficult to index relative abundance of red snapper, no matter what sample sizes are employed.

## 4 Discussion

Simulation results have suggested that the power to detect population trends is typically low for short time series (e.g., 3 years) but is markedly higher for longer time series (e.g., 10 years). Sample sizes seemed less important than increasing the length of the time series in the Gulf; in contrast, increasing sample sizes for the MARMAP survey appeared to be quite important in the south Atlantic. However, using current sampling gears, detecting population trends for red snapper in the south Atlantic from MARMAP data (even with a doubling of sample sizes and a redistribution of effort) still seems elusive, with annual index CVs much higher (e.g.,  $>2.0$ ) than those actually used in stock assessments. Further, an expansion of MARMAP sampling will likely result in age composition samples that are still too low to be practically useful for assessment modeling, at least until the red snapper population starts to recover. Power analyses using MARMAP samples should be taken with a degree of skepticism, however. Future MARMAP sampling will likely employ a combination of video and physical traps, with increased sampling effort in northern Florida and a more comprehensive sampling frame. Thus, this analysis should be revisited after pilot data from a revised survey are collected.

It is worth noting that the power to detect population changes from fishery independent sampling programs is not the same thing as the power to detect population changes from an assessment model. Relative abundance indices are but one piece of a larger jigsaw puzzle that includes age and length compositions, landings, life history data, and other fishery dependent data streams. As such, the power of assessment models to detect population changes may be much higher than power from a fishery independent index alone.

Investigation of abundance metrics from camera trapping data revealed that the MinCount statistic led to biased estimates of population trend, but that it also yielded higher precision than other metrics. Decisions on what metric to use should ultimately be based on whether or not investigators perceive the simulated levels of bias to be biologically meaningful, as well as the increased effort needed to employ other metrics. For instance, the MeanCount metric is unbiased and has a similar CV to MinCount, but the version implemented in simulations requires knowledge of the number of species present on each video frame. In practice, it may be more practical to sample video frames to calculate MeanCount (e.g., by reading every  $n$ th frame).

The presence/absence metric is nonlinearly related to abundance if taken at face

value, but may prove useful if the relationship between occupancy and abundance is explicitly modeled. For instance, recent advancements in occupancy modeling have received widespread attention (e.g., MacKenzie et al., 2002; Royle and Nichols, 2003), particularly in research on terrestrial species. These methods assume that a species is present in an area if it is detected, but that an observed absence cannot be equated to true absence because of the possibility that it is present but not detected. In order to make robust inferences about true occupancy, one must collect data that can be used to estimate detection probabilities, as with repeat site visits. By exploiting the relationship between detection probability and abundance noted by Royle and Nichols (2003; Eq. 1), one can specify a hierarchical model for abundance that is conditional only on presence/absence data (where some of the sites are visited for than once). For example, consider the following model:

$$\begin{aligned} N_{it} &\sim \text{Poisson}(\lambda_{it}), \\ p_{it} &= 1 - (1 - r_{it})^{N_{it}}, \\ X_{it} &\sim \text{Bernoulli}(p_{it}). \end{aligned}$$

Here,  $X_{it} = 1$  if the focal species is observed during visit  $t$  to site  $i$ , and is zero otherwise. The intensity of the Poisson process ( $\lambda$ ) could be written as a function of site specific covariates and as a function of time (denoted by the subscript  $t$ ); if needed, detection probabilities  $r_i$  could also be parameterized in terms of site specific covariates. One could also contemplate an alternative to the Poisson distribution (e.g., zero inflated Poisson, negative binomial) in order to address the inevitable overdispersion observed in fisheries data.

Royle (2004) suggested another approach that might be used to account for detection probabilities when estimating relative abundance from count data. Originally developed for avian point counts, Royle’s approach involves modeling spatially replicated counts using an  $N$ -mixture model. For instance, one can conceptualize the true abundance at a specific sampling location as arising from some underlying mixture distribution (e.g., Poisson). The spatial replication is sufficient to permit estimation of a detection parameter, and standard frequentist inference can be performed by integrating out the unobserved abundance parameters from the likelihood.

## 5 Acknowledgments

The author is grateful for helpful comments and suggestions by N. Bacheler, J. Walter, and E. Williams on an earlier draft; M. Reichert, J. Stephen for providing MARMAP

data; and C. Gledhill for providing SEAMAP data.

## References

- Czado, C., V. Erhardt, A. Min, and S. Wagner. 2007. Zero-inflated generalized Poisson models with regression effects on the mean, dispersion, and zero-inflation level applied to patent outsourcing rates. *Statistical Modelling* **7**:125–153.
- Ellis, D., and E. DeMartini. 1995. Evaluation of a video camera technique for indexing abundance of juvenile pink snapper, *Pristipomoides filamentosus*, and other Hawaiian insular shelf fishes. *Fisheries Bulletin* **93**:67–77.
- Lo, N., L. Jacobson, and J. Squire. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. *Canadian Journal of Fisheries and Aquatic Sciences* **49**:2515–2526.
- MacKenzie, D., J. Nichols, G. Lachman, S. Droege, J. Royle, and C. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**:2248–2255.
- R Development Core Team, 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Renshaw, E. 1991. Modelling biological populations in space and time. Cambridge University Press, Cambridge, UK.
- Royle, J. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**:108–115.
- Royle, J., and J. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology* **84**:777–790.
- Searle, S., F. Speed, and G. Milliken. 1980. Population marginal means in the linear model: An alternative to least squares means. *American Statistician* **34**:216–221.
- Seber, G. A. F. 1982. *The Estimation of Animal Abundance and Related Parameters*, 2nd Edition. Blackburn Press, Caldwell, NJ, USA.
- Willis, T., R. Millar, and R. Babcock. 2000. Detection of spatial variability in relative density of fishes: comparison of visual census, angling, and baited underwater video. *Marine Ecology Progress Series* **198**:249–260.

Figure 1: A map giving locations of first stage sampling cells for Gulf of Mexico SEAMAP reef fish video survey. Reproduced from "An Internal Review of the SEFSC Resource Surveys Program" (Annex Figure B).

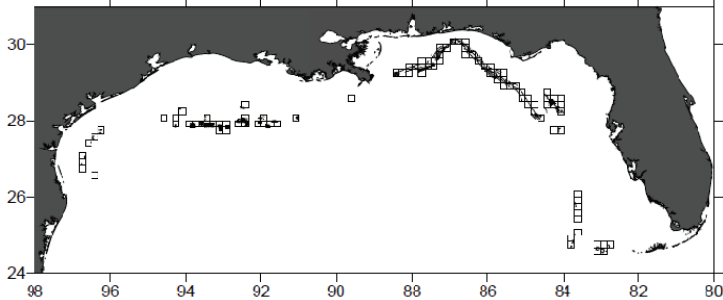


Figure 2: Relationship between true abundance at a given site ( $N$ ) and the probability of observing the focal species at a given site as a function of individual detection probabilities ( $r$ ). The relationship is nonlinear, suggesting the use of raw % presence will perform poorly as a relative abundance index.

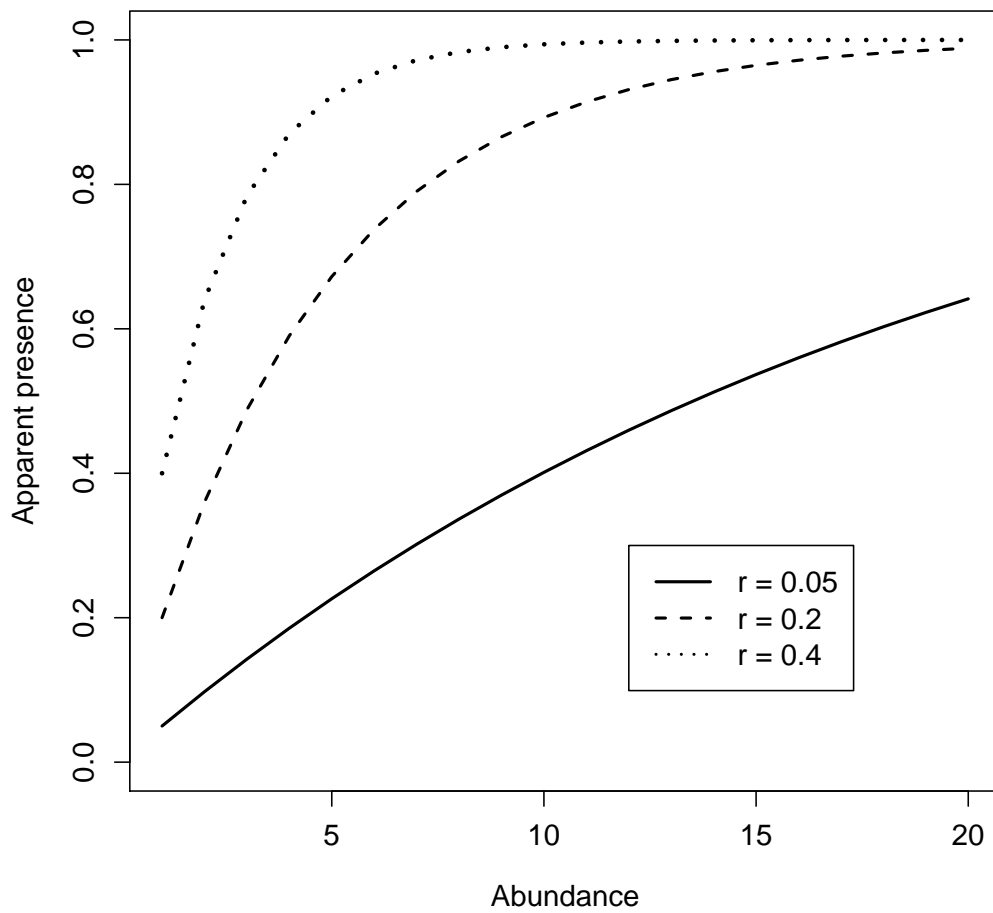


Figure 3: Relative abundance trends as calculated by several underwater camera metrics on simulated data. Each point represents the average relative abundance value for 1000 simulations with an assumed true abundance given by the x-axis. Each simulation depended on  $\alpha_1$ , the rate at which individual fish enter the camera's field of vision, and  $\alpha_2$ , the rate at which individual fish leave the camera's field of vision. The number of fish in the camera's field of vision at the start of sampling sampled from a Binomial( $N_i, 0.1$ ) distribution in all simulations. Panels on the left give results for  $\alpha_1 = 0.05/\text{minute}, \alpha_2 = 0.2/\text{minute}$ , while panels on the right give results for  $\alpha_1 = 0.5, \alpha_2 = 0.5$ . Each simulation replicated a 20 minute time series. True abundance is provided by the dashed line; the relative abundance series was standardized to the first data point to facilitate comparison.

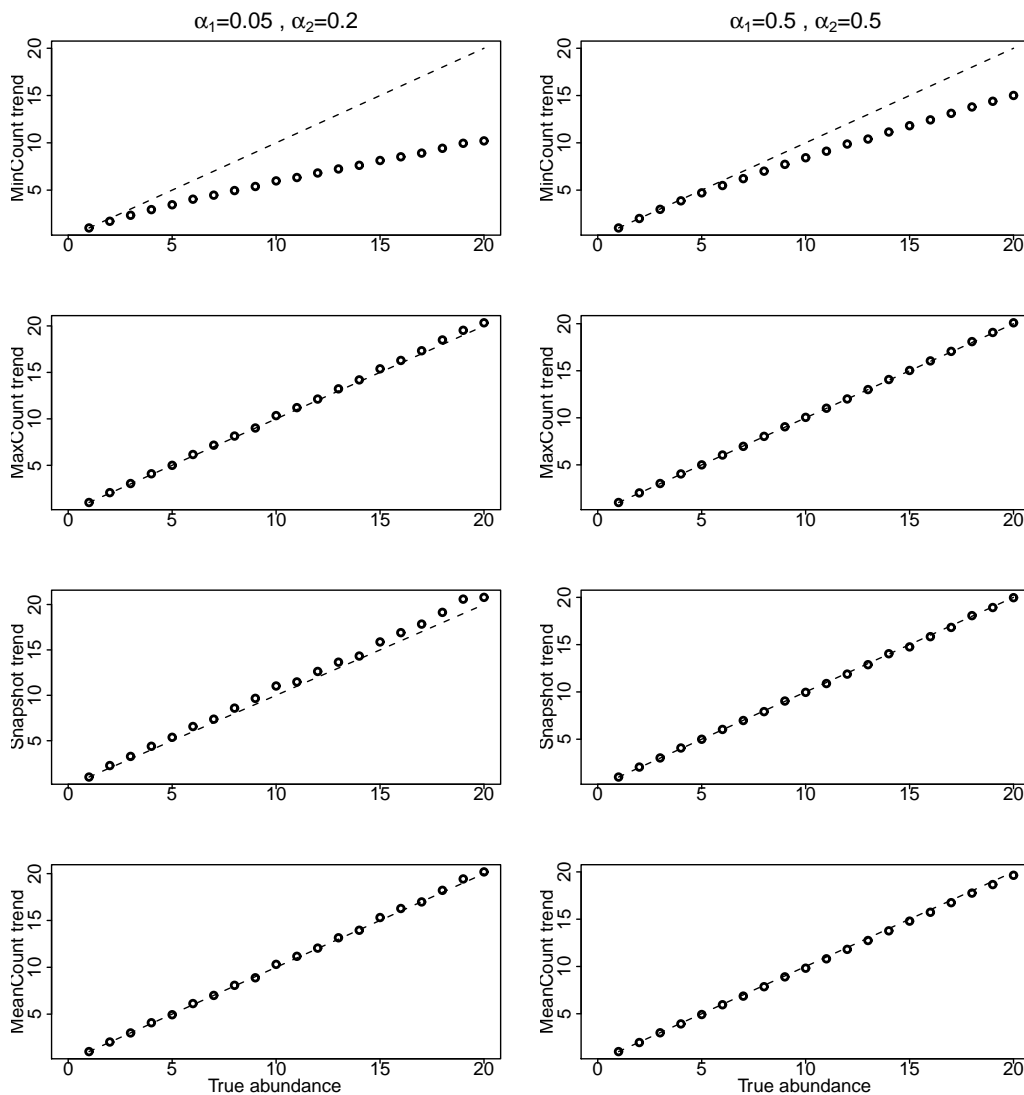


Figure 4: Coefficient of variation for different underwater camera metrics as computed from simulated data. Each point represents the CV for 1000 simulations at a given value of true abundance. Each simulation depended on  $\alpha_1$ , the rate at which individual fish enter the camera's field of vision, and  $\alpha_2$ , the rate at which individual fish leave the camera's field of vision. The number of fish in the camera's field of vision at the start of sampling sampled from a  $\text{Binomial}(N_i, 0.1)$  distribution in all simulations. Each simulation replicated a 20 minute time series.

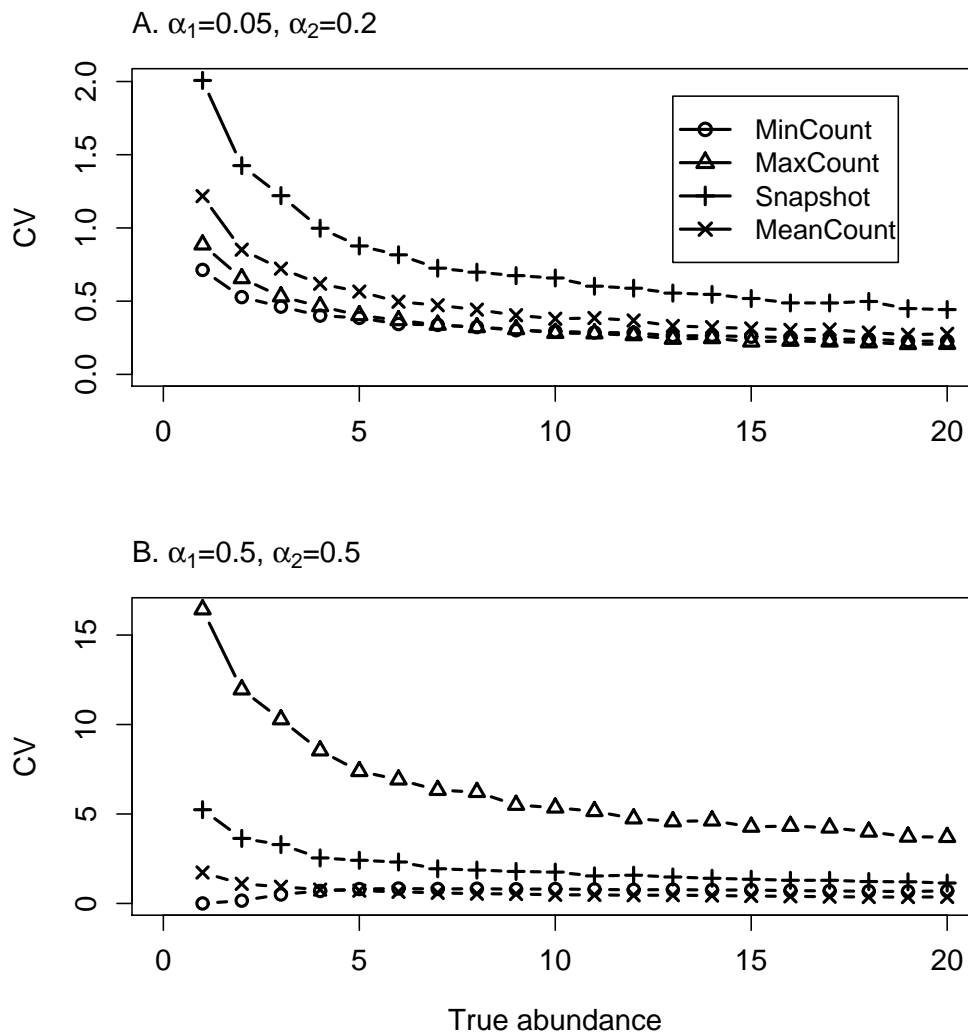




Figure 5: Normal distributions (solid line) fit to estimated strata-specific Poisson intensity parameters (points) for SEAMAP reef fish surveys in the Gulf of Mexico. Estimated distributions were used to sample strata means for unsampled strata in the simulation study, with the constraint that  $\mu_s > 0$ .

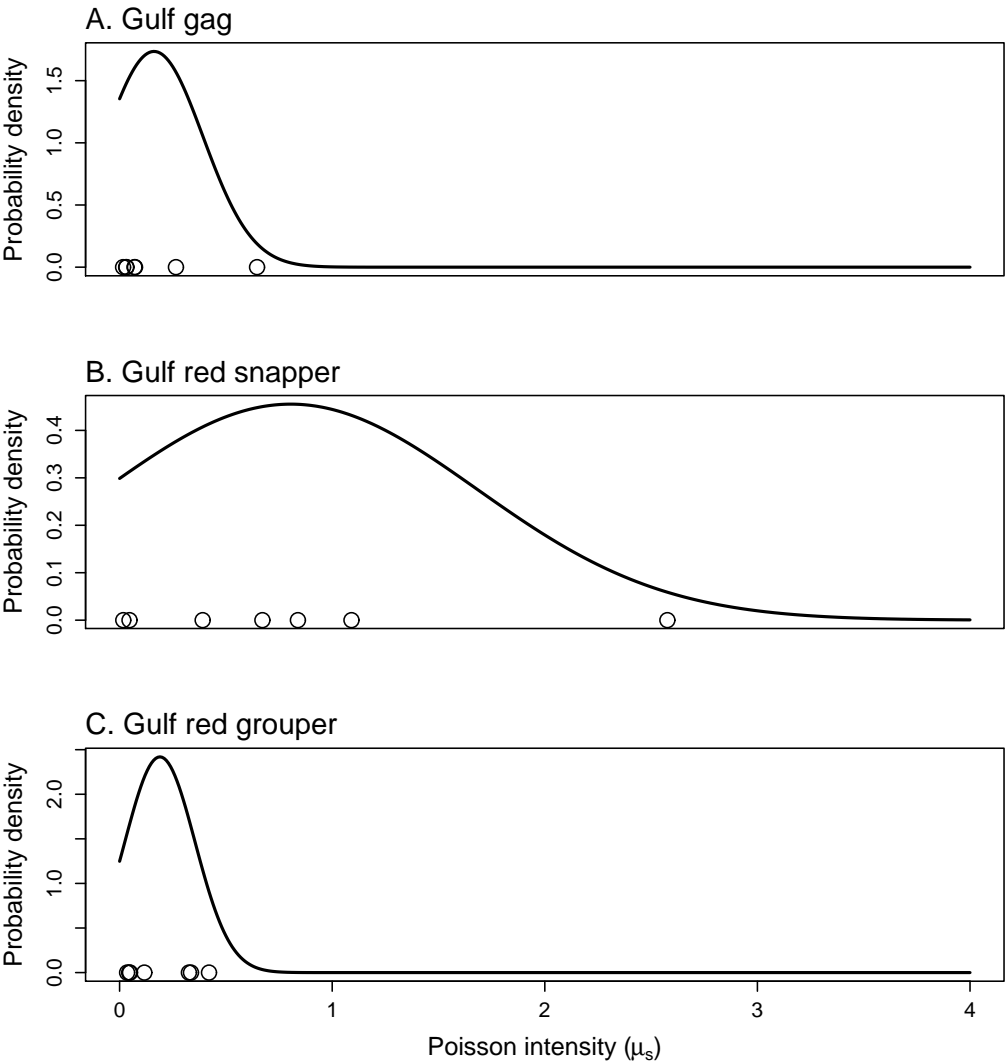


Figure 6: Power to detect population trends for gag grouper from SEAMAP reef fish survey in the Gulf of Mexico. The heavy solid line gives power when annual sample sizes at current levels, while while heavy dashed and heavy dotted lines give power for half and twice the current levels of sampling, respectively. Each column of plots gives results for different lengths of time series (3 years, 5 years, 10 years). Each row of plots is associated with a different simulation scenario (e.g., top plots give results for the case where sample coverage is 50% and there is no heterogeneity in population trend among spatial strata; the final row gives results for the case where sample coverage is 100% and where spatial strata have different relative abundance trends). The thin dashed lines give 95% confidence limits for the nominal  $\alpha$  level (nominal  $\alpha$  is 0.1); thin solid lines give realized  $\alpha$  levels from simulations set so as to have no overall trend.

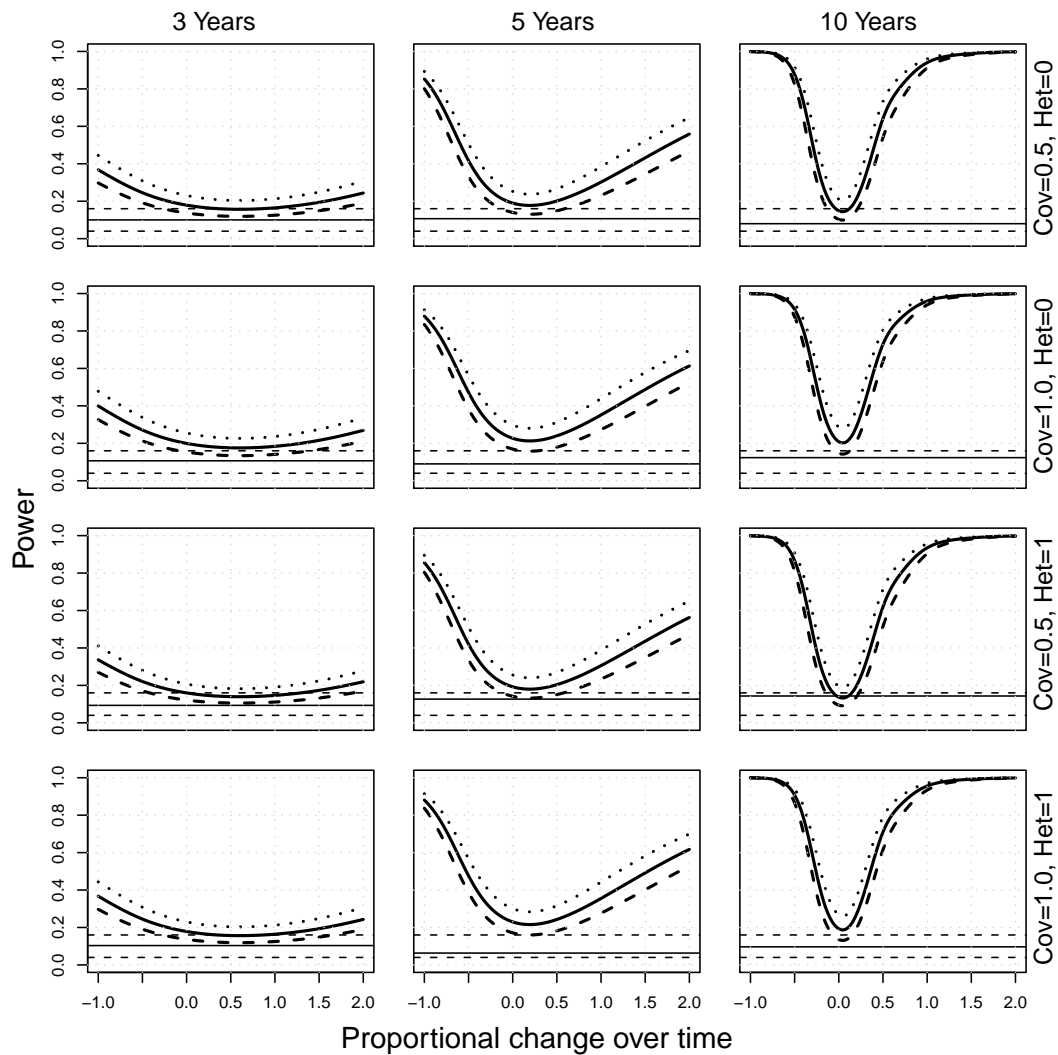


Figure 7: Power to detect population trends for red grouper from SEAMAP reef fish survey in the Gulf of Mexico. The heavy solid line gives power when annual sample sizes at current levels, while heavy dashed and heavy dotted lines give power for half and twice the current levels of sampling, respectively. Each column of plots gives results for different lengths of time series (3 years, 5 years, 10 years). Each row of plots is associated with a different simulation scenario (e.g., top plots give results for the case where sample coverage is 50% and there is no heterogeneity in population trend among spatial strata; the final row gives results for the case where sample coverage is 100% and where spatial strata have different relative abundance trends). The thin dashed lines give 95% confidence limits for the nominal  $\alpha$  level (nominal  $\alpha$  is 0.1); thin solid lines give realized  $\alpha$  levels from simulations set so as to have no overall trend.

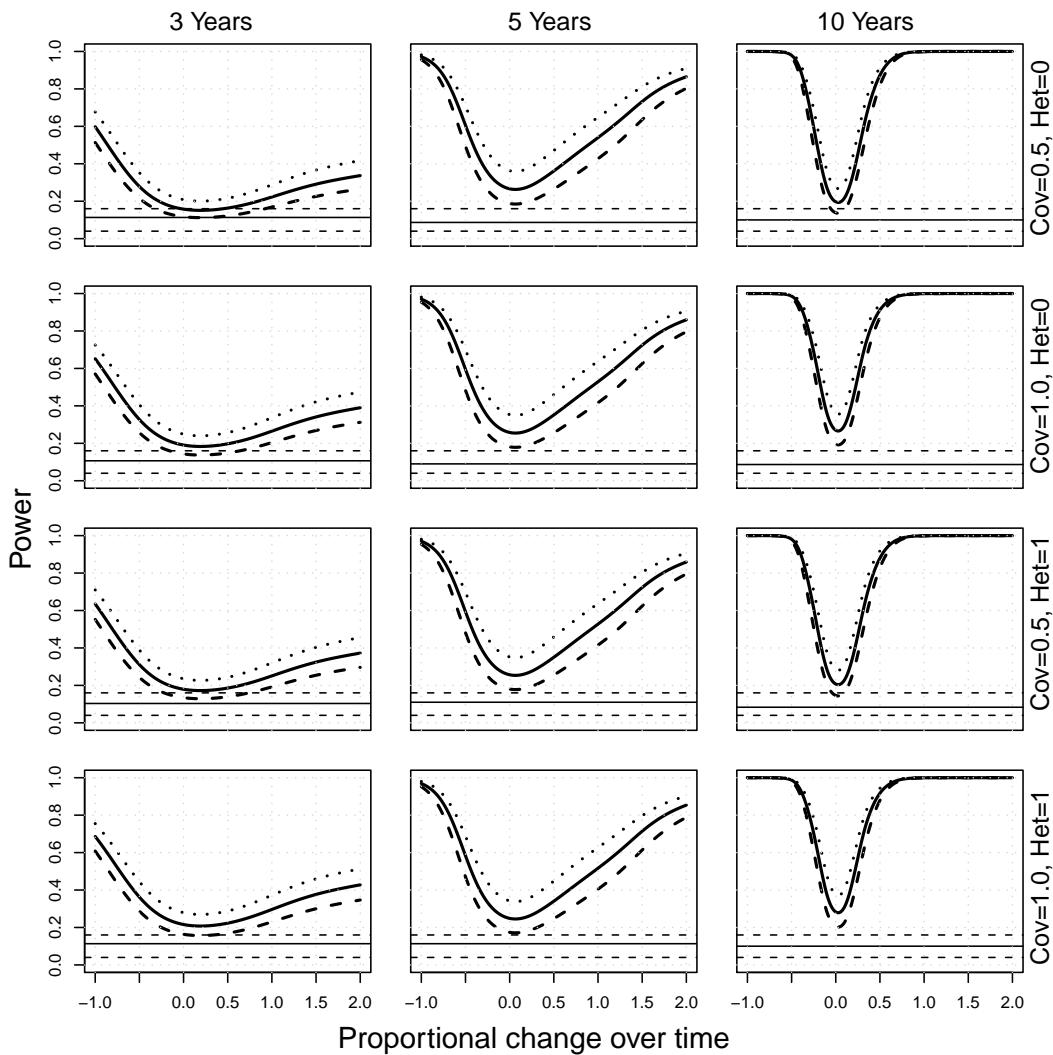


Figure 8: Power to detect population trends for red snapper from SEAMAP reef fish survey in the Gulf of Mexico. The heavy solid line gives power when annual sample sizes at current levels, while while heavy dashed and heavy dotted lines give power for half and twice the current levels of sampling, respectively. Each column of plots gives results for different lengths of time series (3 years, 5 years, 10 years). Each row of plots is associated with a different simulation scenario (e.g., top plots give results for the case where sample coverage is 50% and there is no heterogeneity in population trend among spatial strata; the final row gives results for the case where sample coverage is 100% and where spatial strata have different relative abundance trends). The thin dashed lines give 95% confidence limits for the nominal  $\alpha$  level (nominal  $\alpha$  is 0.1); thin solid lines give realized  $\alpha$  levels from simulations set so as to have no overall trend.

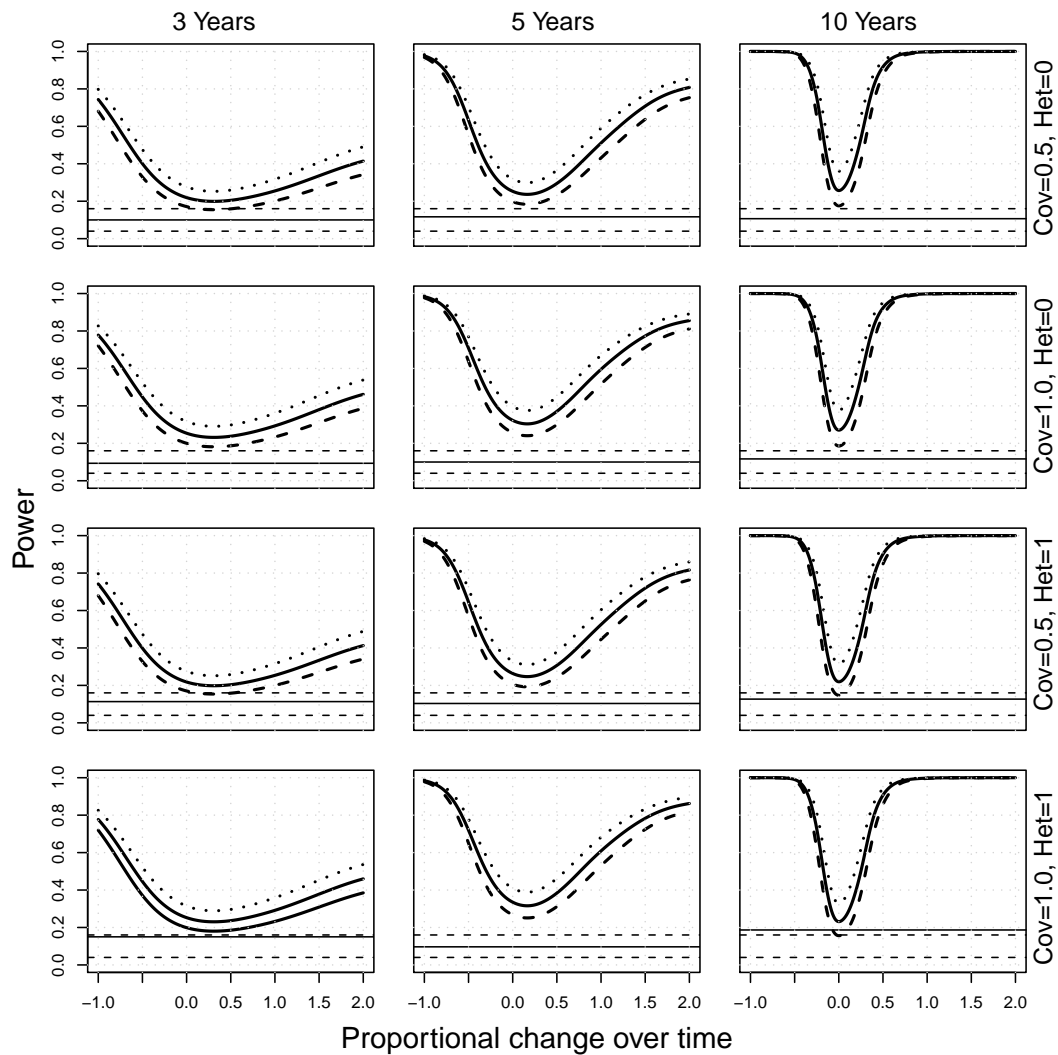


Figure 9: Effects of several explanatory variables on the Poisson intensity parameter for black sea bass catch per set from MARMAP chevron traps in the south Atlantic. All variables were modeled simultaneously within a GAM that used smoothing splines with four degrees of freedom for each variable (with the exception of the Year effect, which had six). The relative sample size at each value of the explanatory variables is shown along the x-axis.

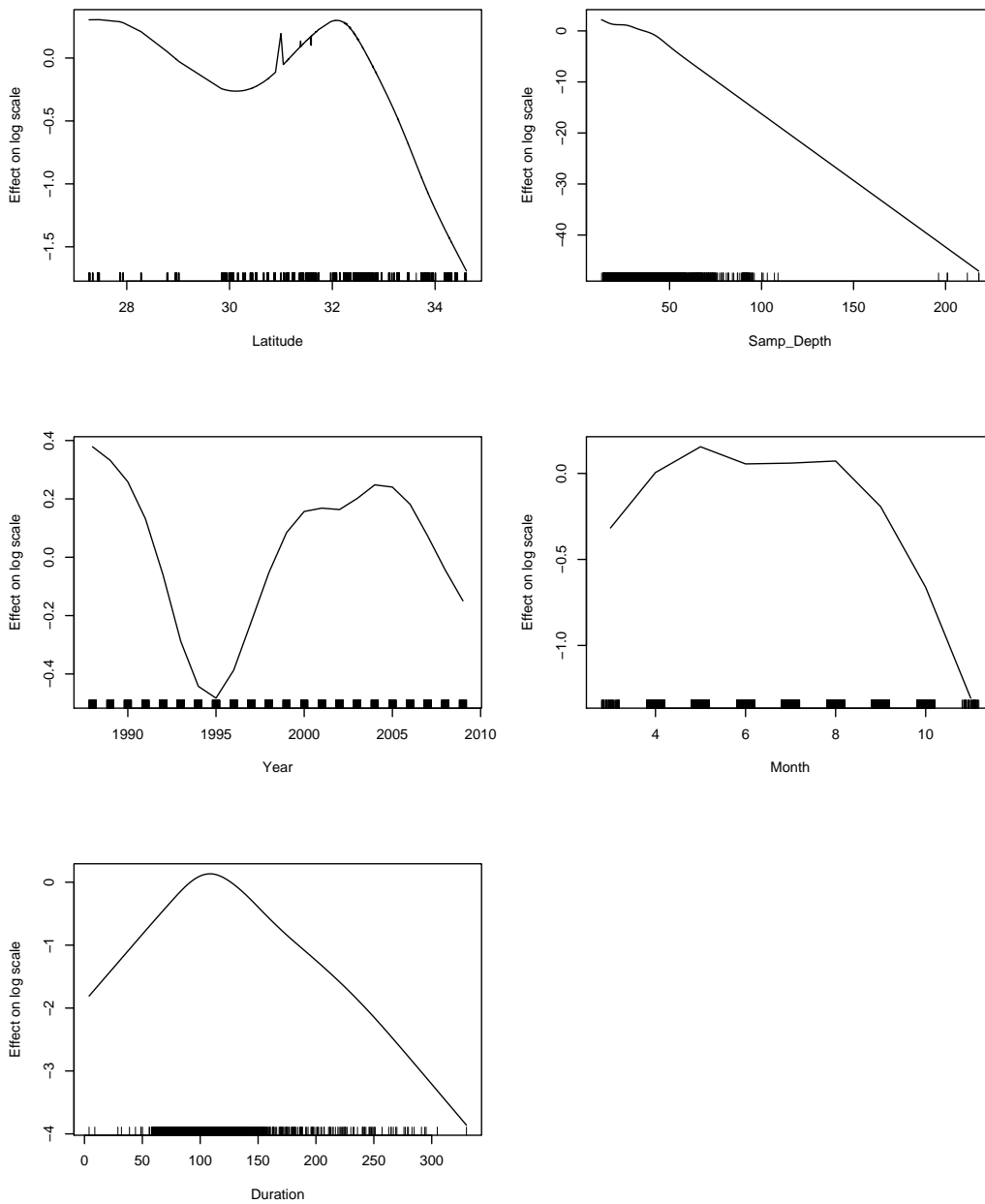


Figure 10: Effects of several explanatory variables on the Poisson intensity parameter for red porgy catch per set from MARMAP chevron traps in the south Atlantic. All variables were modeled simultaneously within a GAM that used smoothing splines with four degrees of freedom for each variable (with the exception of the Year effect, which had six). The relative sample size at each value of the explanatory variables is shown along the x-axis.

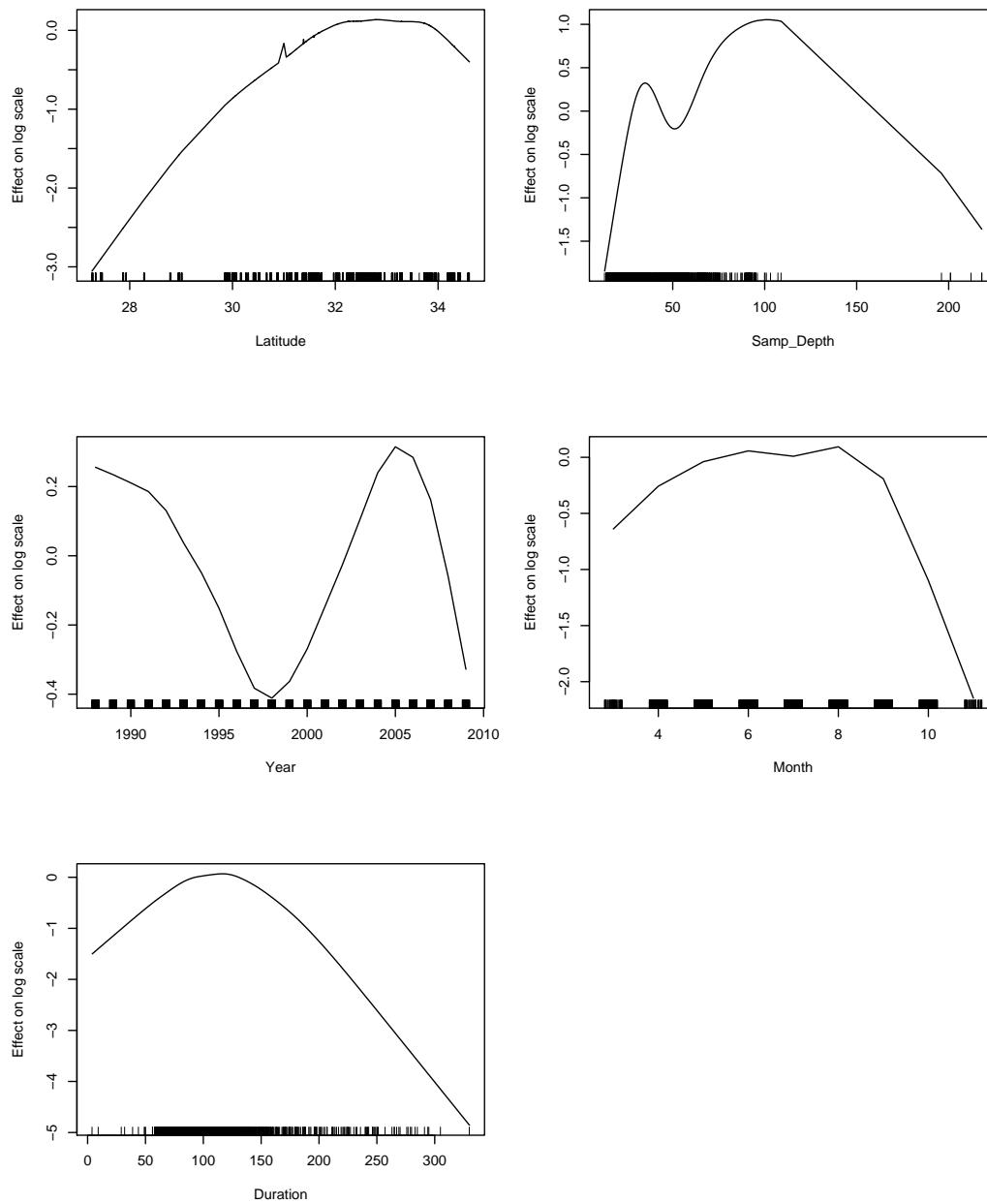


Figure 11: Effects of several explanatory variables on the Poisson intensity parameter for red snapper catch per set from MARMAP chevron traps in the south Atlantic. All variables were modeled simultaneously within a GAM that used smoothing splines with four degrees of freedom for each variable (with the exception of the Year effect, which had six). The relative sample size at each value of the explanatory variables is shown along the x-axis.

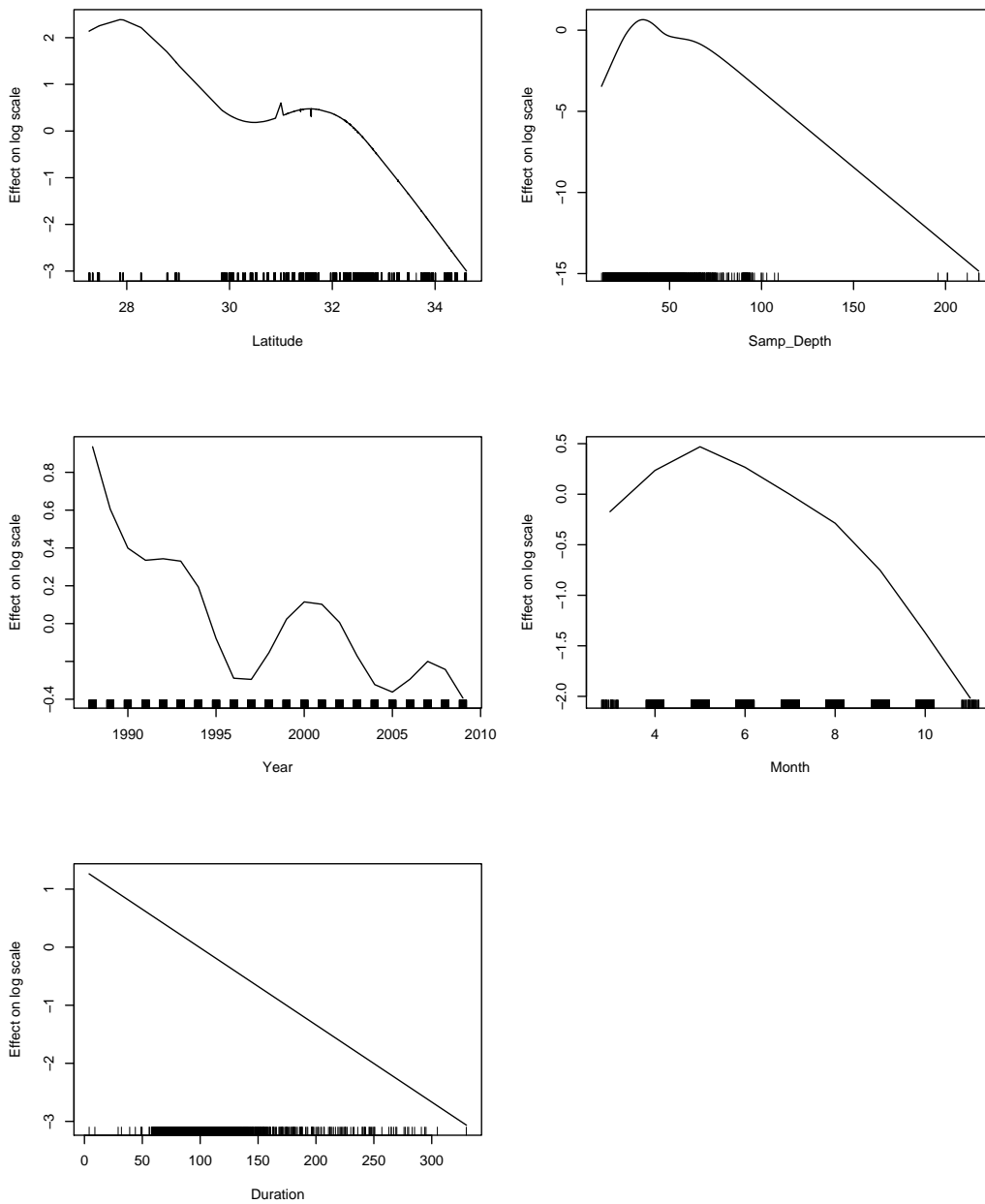


Figure 12: A histogram showing the frequency of MARMAP Chevron trap sets by latitude (degrees north), 1988-2009.

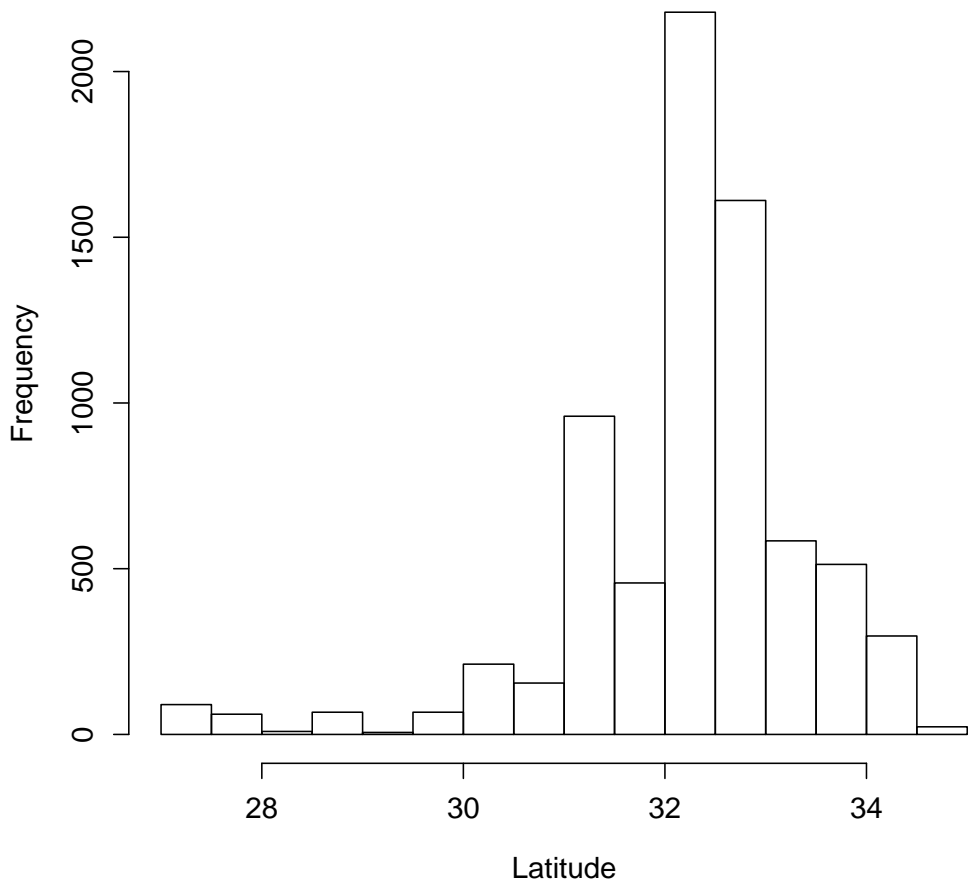




Figure 13: Power to detect population trends for black sea bass in the south Atlantic based on four simulated scenarios. The top left panel corresponds to current levels of MARMAP sampling ('Status quo'); the top right panel corresponds to a range expansion ('Expansion 1'); bottom left corresponds to a range expansion with a doubling of sampling effort ('Expansion 1+'); bottom right shows results for a range expansion, doubling of sampling effort, and disproportionately high sampling in north Florida ('Expansion 2'). The solid line gives power as a function of proportional population change for a three year time series, while the dashed line represents a five year time series and the dotted line represents a ten year time series.

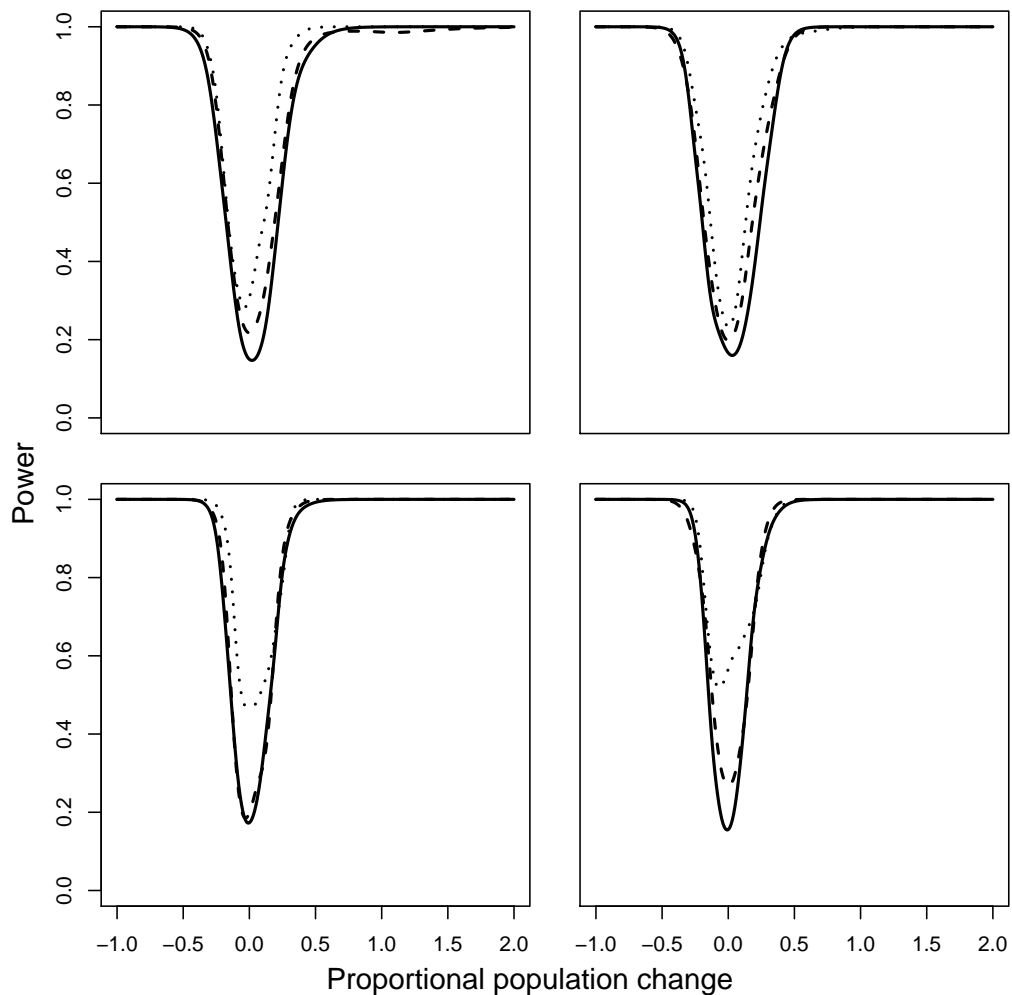


Figure 14: Power to detect population trends for red porgy in the south Atlantic based on four simulated scenarios. The top left panel corresponds to current levels of MARMAP sampling ('Status quo'); the top right panel corresponds to a range expansion ('Expansion 1'); bottom left corresponds to a range expansion with a doubling of sampling effort ('Expansion 1+'); bottom right shows results for a range expansion, doubling of sampling effort, and disproportionately high sampling in north Florida ('Expansion 2'). The solid line gives power as a function of proportional population change for a three year time series, while the dashed line represents a five year time series and the dotted line represents a ten year time series.

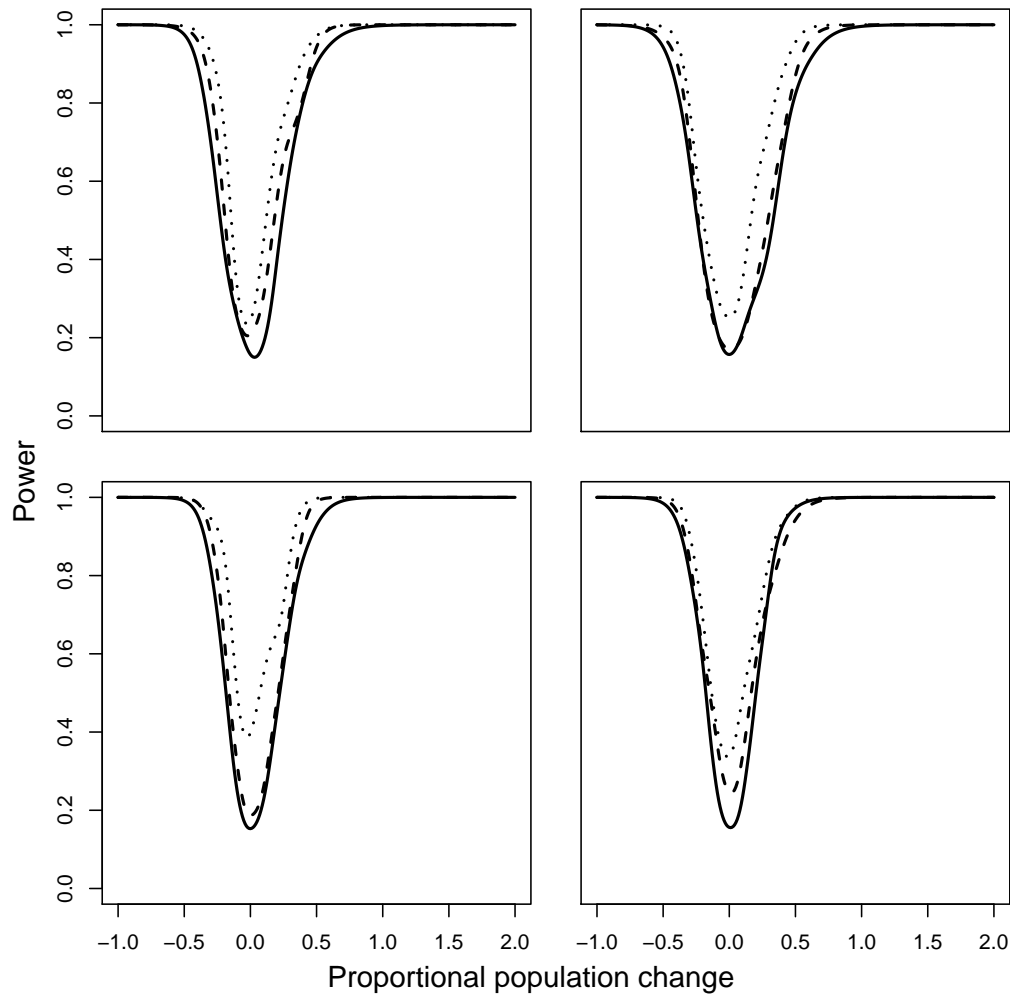


Figure 15: Power to detect population trends for red snapper in the south Atlantic based on four simulated scenarios. The top left panel corresponds to current levels of MARMAP sampling ('Status quo'); the top right panel corresponds to a range expansion ('Expansion 1'); bottom left corresponds to a range expansion with a doubling of sampling effort ('Expansion 1+'); bottom right shows results for a range expansion, doubling of sampling effort, and disproportionately high sampling in north Florida ('Expansion 2'). The solid line gives power as a function of proportional population change for a three year time series, while the dashed line represents a five year time series and the dotted line represents a ten year time series. Predicted power for three year time series is poorly estimated in the tails (most simulated three year time series did not exhibit large changes in abundance).

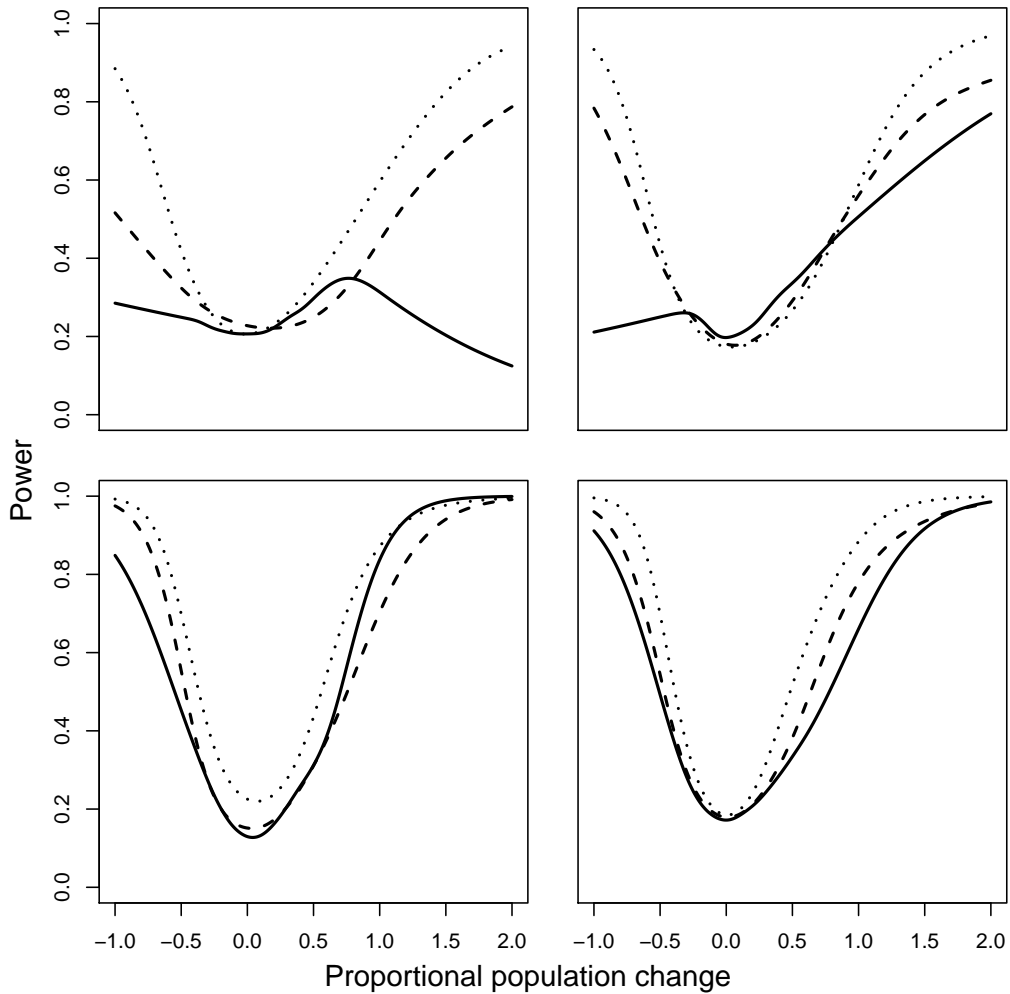


Figure 16: Coefficient of variation for black sea bass and red porgy in the south Atlantic as a function of sample size (number of MARMAP sets) for the 'Expansion 2' scenario (range expansion with disproportionately high sampling in north Florida). The vertical line represents the approximate annual sample sizes currently being collected by the MARMAP survey.

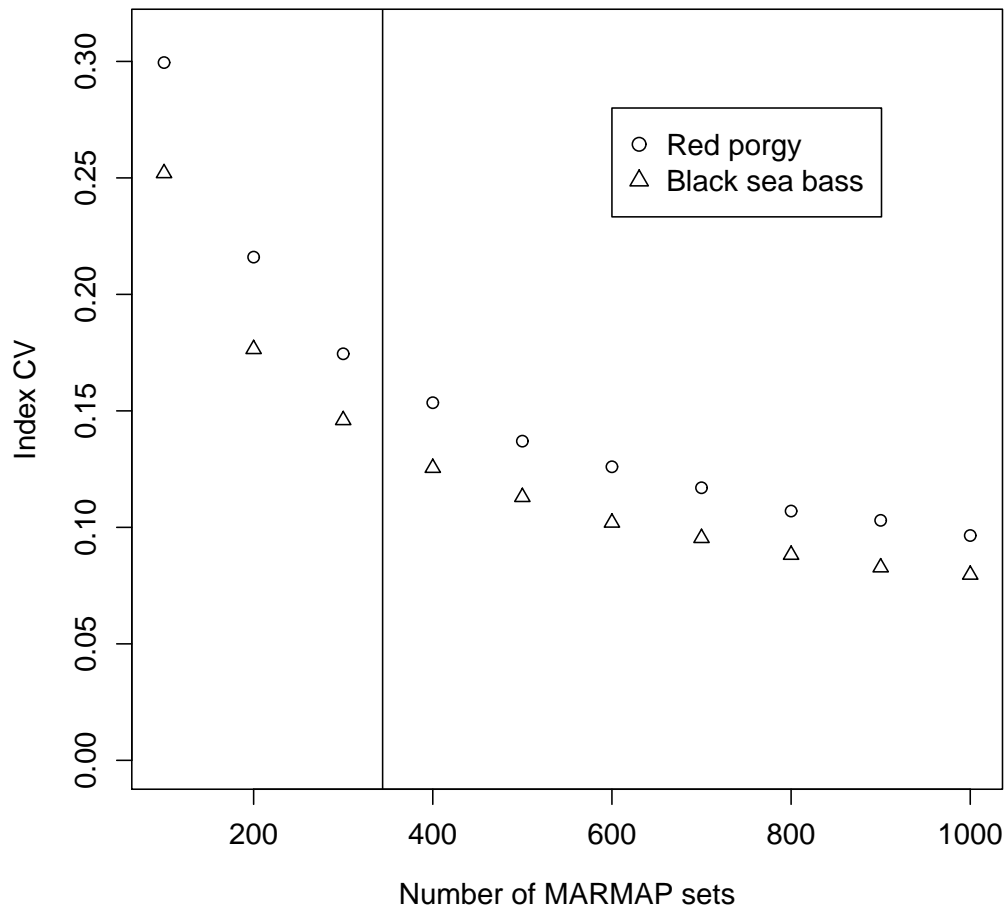


Figure 17: Coefficient of variation for red snapper in the south Atlantic as a function of sample size (number of MARMAP sets) for the 'Expansion 2' scenario (range expansion with disproportionately high sampling in north Florida) (top panel). The vertical line represents the approximate annual sample sizes currently being collected by the MARMAP survey. Also presented are the proportion of simulations that did not converge (bottom panel)

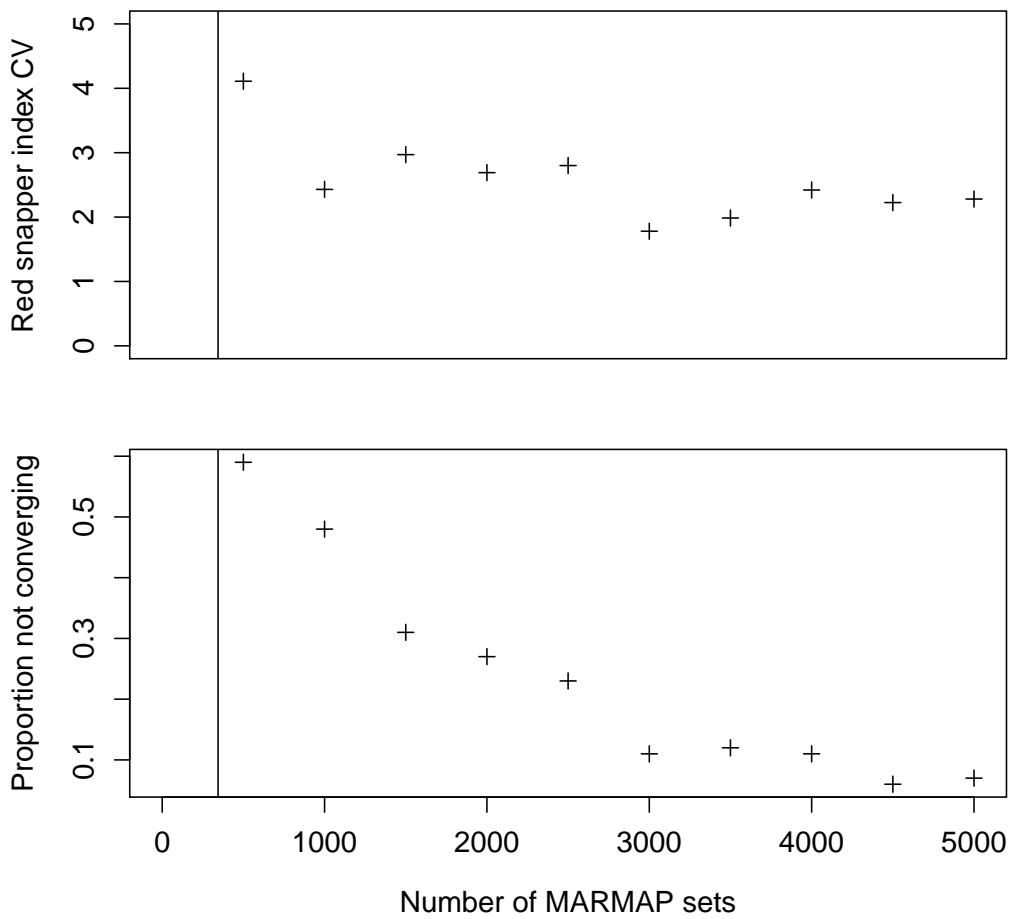


Table 1: Parameter estimates (on the real scale) from fitting ZIGP models to SEAMAP reef fish survey data. Estimates of strata-specific  $\mu$  parameters were used as starting values for the first year of simulations; estimated variance inflation values ( $\phi$ ) and zero-inflation factors ( $\omega$ ) were treated as constants throughout simulations.

Parameter	Species		
	Gag	Red grouper	Red snapper
$\mu_1$	0.03	0.12	0.05
$\mu_2$	0.03	0.33	0.02
$\mu_3$	0.65	0.42	0.67
$\mu_4$	0.27	0.34	0.39
$\mu_5$	0.07	0.04	0.84
$\mu_6$	0.07	0.05	1.09
$\mu_7$	0.02	0.04	2.57
$\omega$	0.38	0.00	0.19
$\phi$	1.24	1.09	2.32