ESTIMATING FISH POPULATIONS FROM REEF CITIZEN SCIENCE VOLUNTEER DIVER ORDER-OF-MAGNITUDE SURVEYS

JOHN R. WOLFE
Advanced Assessment Team Volunteer
Reef Environmental Education Foundation (REEF)
2320 Blake Street
Berkeley, CA 94704
john.wolfe@tippingmar.com

CHRISTY V. PATTENGILL-SEMMENS
Director of Science
Reef Environmental Education Foundation (REEF)
PO Box 370246
Key Largo, FL 33037
christy@REEF.org

ABSTRACT

We describe several models to convert order-of-magnitude count data to a numeric mean and demonstrate that with a sufficient number of surveys, estimates of the mean with a reasonably small confidence interval can be attained. The method is applied to fish survey data collected as part of the Reef Environmental Education Foundation (REEF) Volunteer Survey Project, a citizen science program that has accumulated a database of over 172,000 surveys by recreational divers using the Roving Diver Technique (RDT). We compare three models to convert RDT order-of-magnitude counts to expected arithmetic means. For each model, parameter estimates and associated confidence intervals were derived from 292 RDT surveys where precise counts were also made. Models were compared using the small sample Akaike Information Criteria (AICc). The best-fitting model uses disaggregated bin-count data and considers the relative proportion of counts in adjacent bins.

INTRODUCTION

There is growing interest by scientists, resource managers, and decision makers in the potential role of data collected by citizen scientists as another source for tracking fish populations. Such data sets may be applied to stock assessments, the evaluation of California's network of Marine Protected Areas (MPAs), and how marine fishes respond to changing ocean conditions. Moreover, the species of nearshore rocky reef fishes addressed in this study play an important role in commercial and recreational fisheries, as well as dive tourism. This paper proposes a method to quantitatively analyze data like that generated by the Reef Environmental Education Foundation (REEF) Fish Survey Project, a citizen science data collection program conducted by recreational divers.

Because biological populations tend to grow and decline exponentially, population densities often vary across both time and space by orders of magnitude (May 1975; Engen and Lande 1996). One of the most efficient methods of surveying such populations is with order-of-magnitude counting methods. REEF developed the Roving Diver Technique (RDT) to enlist SCUBA divers to conduct fish surveys (Semmens et al. 2000; Schmitt et al. 2002). The surveyors roam across a dive site and record order-of-magnitude counts of fish species they observe and can positively identify, with the following counting bins: **S**ingle = 1, **F**ew = 2–10, **M**any = 11–100, and **A**bundant >100 (hereafter termed SFMA data). As of July 2013, over 172,000 surveys have been made worldwide, with the results made publicly available on REEF's website, www.REEF.org. In the Monterey Peninsula area of the California coast, 3,157 surveys were conducted over fifteen years, from 1997 through 2011. These data, collected by volunteer citizen science divers, have great potential to augment and strengthen regional scientific, conservation, and management efforts (Wolfe and Pattengill 2013; Holt et al. 2013). This paper identifies computational techniques to convert REEF categorical data to arithmetic means, thereby enhancing its statistical usefulness.

Our results presented here show that, with a sufficient number of surveys, order-of-magnitude count data can be converted to numeric means with reasonably small confidence intervals. Because RDT dives have mean durations covering mean distances, the average number of fish seen per dive can be used to approximate relative fish density. We base our analysis on 292 RDT surveys where exact counts were made instead of order-of-magnitude SFMA counts. The exact count data were converted to SFMA data for 36 species, and then three different models were compared to determine which most accurately converts SFMA data back to arithmetic means. Akaike Information Criterion for small samples (AICc) was used to determine the model that gave the best, most parsimonious fit. While all three estimation models were compared using the same SFMA data, the first model aggregated the SFMA data into a single log-density score *DEN*, while the second and third models were based on the disaggregated SFMA bin data. The latter two models proved to be more precise, with significantly tighter confidence intervals around the estimated mean. We then quantified the confidence interval associated with the best estimation method and examined other sources of uncertainty.

TABLE 1
SFMA counts and numeric mean for 36 fish species based on 292 exact-count surveys. Values under columns S, F, M, and A are the number of dives for which that species was recorded in that $\log_{10}$ counting bin.

	Species Name		1 S	2–10 F	11–100 M	101+ A	Total	Obs. Mean
1	Striped Seaperch	*Embiotoca lateralis*	6	90	167	1	264	17.1
2	Blue Rockfish	*Sebastes mystinus*	5	41	121	91	258	114.6
3	Kelp Rockfish	*Sebastes atrovirens*	12	133	106	0	251	12.7
4	Kelp Greenling	*Hexagrammos decagrammus*	56	171	4	0	231	3.3
5	Pile Perch	*Damalichthys vacca*	30	148	41	1	220	8.0
6	Painted Greenling	*Oxylebius pictus*	36	140	31	0	207	6.3
7	Black Perch	*Embiotoca jacksoni*	40	133	20	0	193	4.6
8	Black & Yellow Rockfish	*Sebastes chrysomelas*	47	138	6	0	191	3.4
9	Blackeye Goby	*Coryphopterus nicholsi*	28	116	24	1	169	6.6
10	Black Rockfish	*Sebastes melanops*	41	107	21	0	169	5.6
11	Senorita	*Oxyjulis californica*	13	29	89	35	166	67.1
12	Lingcod	*Ophiodon elongatus*	72	83	0	0	155	2.0
13	YOY Rockfish	*Sebastes* spp. (<5 cm)	12	30	62	49	153	125.7
14	Tubesnout	*Aulorhychus flavidus*	19	22	50	52	143	143.4
15	Olive Rockfish	*Sebastes serranoides*	41	84	13	0	138	4.1
16	Gopher Rockfish	*Sebastes carnatus*	47	82	3	0	132	3.2
17	Cabezon	*Scorpaenichthys marmoratus*	84	43	0	0	127	1.6
18	Kelp Perch	*Brachyistius frenatus*	36	54	26	2	118	10.3
19	Rubberlip Surfperch	*Rhacochilus toxotes*	26	63	16	0	105	5.6
20	Reef Surfperch	*Micrometrus aurora*	4	25	34	0	63	15.8
21	Rainbow Surfperch	*Hypsurus caryi*	16	25	18	0	59	10.4
22	Copper Rockfish	*Sebastes caurinus*	25	24	2	0	51	2.5
23	Snubnose Sculpin	*Orthonopias triacis*	31	20	0	0	51	1.7
24	Speckled Sanddab	*Citharichthys stigmaeus*	12	32	5	0	49	4.0
25	Yellowtail Rockfish	*Sebastes flavidus*	18	25	4	0	47	3.4
26	Gibbonsia Kelpfish	*Gibbonsia* spp.	39	6	0	0	45	1.2
27	Treefish	*Sebastes serriceps*	28	10	0	0	38	1.4
28	Monkeyface Prickleback Eel	*Cebidichthys violaceus*	27	10	0	0	37	1.4
29	Vermillion Rockfish	*Sebastes miniatus*	32	4	0	0	36	1.1
30	Opaleye	*Girella nigricans*	14	19	2	0	35	2.8
31	California Sheephead	*Semicossyphus pulcher*	14	21	0	0	35	2.2
32	Blacksmith	*Chromis punctipinnis*	10	11	4	8	33	68.0
33	Scalyhead Sculpin	*Artedius harringtoni*	24	9	0	0	33	1.4
34	Grass Rockfish	*Sebastes rastrelliger*	26	6	0	0	32	1.3
35	Kelp/Calico Bass	*Paralabrax clathratus*	22	6	0	0	28	1.3
36	Coralline Sculpin	*Artedius corallinus*	18	5	0	0	23	1.3

Fish species shown in Tables 1 and 2 are listed in descending order of total sightings over 292 surveys.

METHODS

Exact Count Database: To correlate SFMA data with precise counts of fish seen, exact counts of all observed fish species were recorded while concurrently conducting a REEF RDT survey. 292 such dives were made by lead author JW between 2002 and 2012 on nearshore reefs of south Monterey Bay and Carmel Bay, the area identified by the REEF geozone code prefix 4114. Most dives were from shore at traditionally recognized dive sites extending from Del Monte Beach and the Breakwater in south Monterey Bay to Point Lobos State Marine Reserve in south Carmel Bay. Thirty-six fish species were observed on over 8% of the dives, and these formed the basis of the exact count data set used to derive methods of converting SFMA data to arithmetic means (table 1). Observed numeric means were calculated for each species. The data were also categorized by the lower resolution SFMA $\log_{10}$ counting bins, enumerating how many dives a single fish was seen, or a Few (2–10), Many (11–100), or Abundant (over 100).

Verify Underlying Log-Normal Distribution: To formulate reasonable estimation methodologies, the exact count data were examined to determine the underlying distribution of fish sightings. As expected (Limpert et al. 2001), sightings generally followed a log-normal distribution (figs. 1 and 2). The numeric mean $\bar{x}$ of a log-normal distribution with mean μ and standard deviation σ is:

$$\bar{x} = e^{(\mu + \sigma^2/2)} \tag{1}$$

To determine closeness of fit of species' sightings to a perfect log-normal distribution, the ratios of expected to observed numeric means were plotted against the log-transformed mean abundances for the 36 species. The log-normal distribution assumption fit well for most species, but tended to overestimate the numeric mean for abundant schooling fish species (fig. 3).

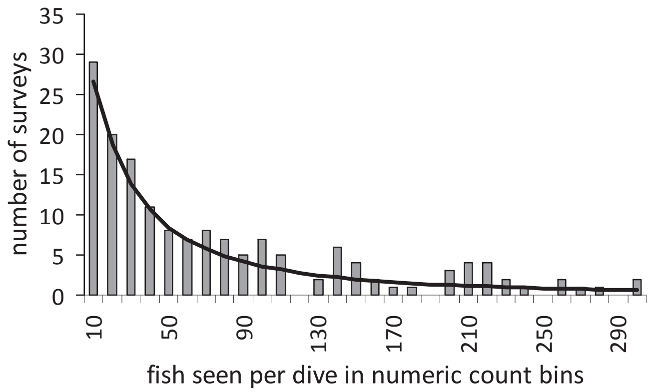

Figure 1a. blue rockfish, numeric x-axis

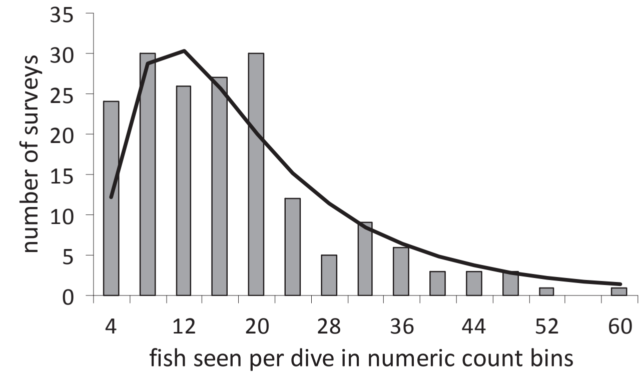

Figure 2a. striped seaperch, numeric x-axis

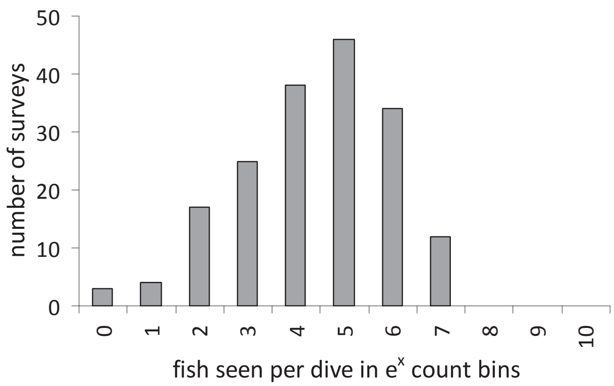

Figure 1b. blue rockfish, logn x-axis

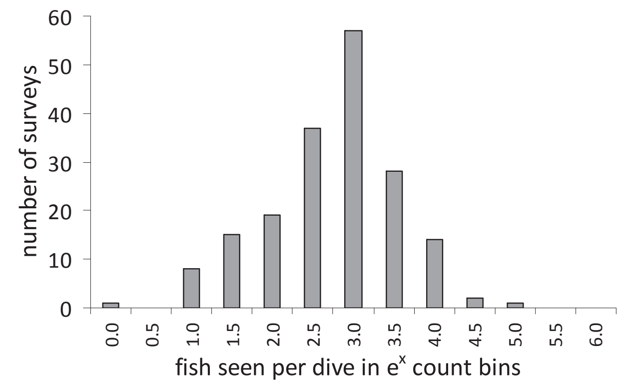

Figure 2b. striped seaperch, logn x-axis

Figures 1 and 2. Figures 1 and 2 refer to blue rockfish and striped seaperch species (left and right graphs), respectively. Figures 1a and 2a (upper graphs) are histograms of the number of surveys (vertical axis) plotted against arithmetic counting bins of fish seen per dive (horizontal axis), with a fitted lognormal distribution (solid line) drawn on top. Figures 1b and 2b (lower graphs) are histograms of the same surveys plotted against natural log counting bins of fish seen per dive.

Because fish counts fit log-normal distributions more closely than normal distributions, Δ used in the least squares regression below were based on proportional rather than arithmetic differences. Likewise, confidence intervals were calculated as "times/divide" $^{x}/$ proportions rather than "plus/minus" $^{+}/_{-}$ arithmetic intervals (Limpert et al. 2001). A times/divide factor of $^{x}/30\%$ (shorthand for x1.30, /1.30) corresponds to plus/minus factors of +30%, –23%.

Application of Least Squares Regression: Three estimation models for converting SFMA data were examined. Optimal parameters for each model were deduced using least squares regression applied to all 36 fish species by calculating the proportional difference, Δ, of the ratio of the expected to observed numeric mean for each species as:

$$\Delta = \ln(ExpectedMean/ObservedMean) \tag{2}$$

For each species, Δ was squared, and the residual sum of the squares (RSS) was calculated by summing Δ^2 for all 36 species, giving each species equal weight in the regression. For each model, optimal parameters to minimize RSS were determined using the Solver Add-On Tool of Microsoft Excel.

Aggregate Log-Density Score, DEN: The REEF Web site reports summaries for groups of surveys for any requested period of time and geographic area within the REEF database. Counts for each species are summarized by sighting frequency (*SF*) and log-density score (*DEN*). The value *DEN* is a single number that aggregates nonzero sightings into a log average:

$$DEN = \frac{S + 2F + 3M + 4A}{S + F + M + A} \tag{3}$$

where:
S = *number of dives reporting a Single individual of a given species*
F = *number of dives reporting a Few (2–10) individuals of that species*
M = *number of dives reporting Many (11–100) individuals of that species*
A = *number of dives reporting Abundant (over 100) individuals of that species*

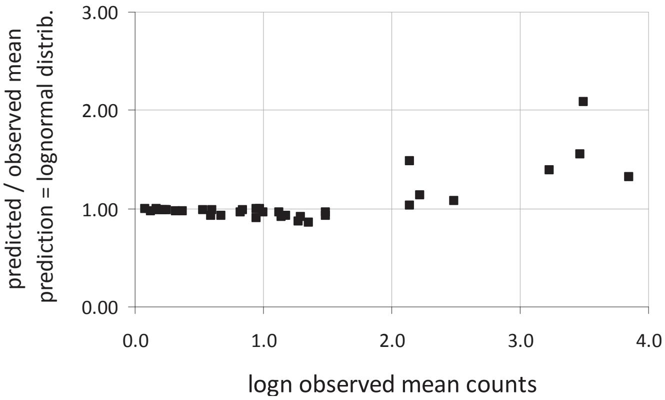

Figure 3. The log-normal distribution's predicted / observed numeric mean ratio plotted as a function of the logn-transformed observed mean counts for a given species; each point represents one of the 36 species observed. Given a perfect log-normal distribution with mean μ and standard deviation σ, predicted numeric mean = exp(μ + σ²/2). The graph suggests that counts of most species closely follow a log-normal distribution, but stray from perfectly log-normal for the most abundant species, where a perfect log-normal distribution predicts higher counts than are actually observed.

Model Descriptions

Models were formulated to predict expected mean from SFMA data. The optimized parameters for these models are intended to apply to all fish species studied, based on best overall fit to 36 fish species reported in the 292 exact count dives. The first model was based on the aggregate log-density score *DEN*, while the second and third models were based on disaggregated SFMA bin counts.

Model 1. The first model used an exponentiation formulation:

$$ExpectedMean_1 = A^{(DEN-1)^B} \tag{4}$$

where A = *a base coefficient* and
B = *an exponent coefficient*.

Model 2. A second model can be devised where a single theoretical average value is sought for each count category, with no consideration of the relative proportion of adjacent count categories. This leads to a simple formulation:

$$ExpectedMean_2 = \frac{S + f_2F + m_2M + a_2A}{S + F + M + A} \tag{5}$$

The parameters f_2, m_2 and a_2 can be regarded as the average number of fish that are seen on dives where 2–10, 11–100, and over 100 fish are seen, respectively.

Model 3. With disaggregated SFMA data, the proportion of sightings in an adjacent category may provide information on the probable average value for the category in question. For instance, for 20 dives with a distribution of 16 Single counts, 4 Few counts, and zero Many and Abundant counts ($S = 16, F = 4, M = 0, A = 0$), it is reasonable to expect the average number of individuals seen in the Few category to be at the lower end of the range of 2 to 10, perhaps between 2 and 3 (fig. 4a). On the other hand, for 20 dives with a distribution of $S = 1, F = 3, M = 12, A = 4$, one would expect a higher average number of individuals seen in the Few category, closer to the upper end of the range 2 to 10 (fig. 4b).

In the hypothetical distributions described, the average number of observed individuals in the "Few" category shifts from 2.4 to 8.4 as the expected mean from all categories increases. An "Average of Few" variable based on the proportion of adjacent count categories can be formulated, as well as similar variables for the Many and Abundant categories. These variables, *AvgF, AvgM* and *AvgA*, are bound by the limits that define their ranges, and are formulated as follows:

$$AvgF = \frac{2S + f_fF + 10M}{S + F + M}$$

$$AvgM = \frac{11F + m_mM + 100A}{F + M + A} \tag{6}$$

$$AvgA = \frac{a_mM + a_aA}{M + A}$$

f_f = *contribution of Few component to AvgF*
m_m = *contribution of Many component to AvgM*
a_m = *contribution of Many component to AvgA*
a_a = *contribution of Abundant component to AvgA*

The variables *AvgF*, *AvgM* and *AvgA*, multiplied by their corresponding category counts *F*, *M* and *A*, can then be summed and divided by the total nonzero counts to give the expected average sightings per dive:

$$ExpectedMean_3 = \frac{S + F \cdot AvgF + M \cdot AvgM + A \cdot AvgA}{S + F + M + A} \tag{7}$$

Incorporation of Zero Counts

In all the models described above, the expected mean represents the mean for nonzero surveys. This nonzero mean is multiplied by the sighting frequency (fraction of nonzero sightings) to calculate an overall average number of fish of a given species seen per dive.

$$ExpectedMean\ (AllSurveys) = SightingFrequency \cdot ExpectedMean(NonZeroSurveys) \tag{8}$$

Akaike Information Criterion (AICc)

To determine best fit while avoiding the pitfalls of either underfitting or overfitting the data, models were

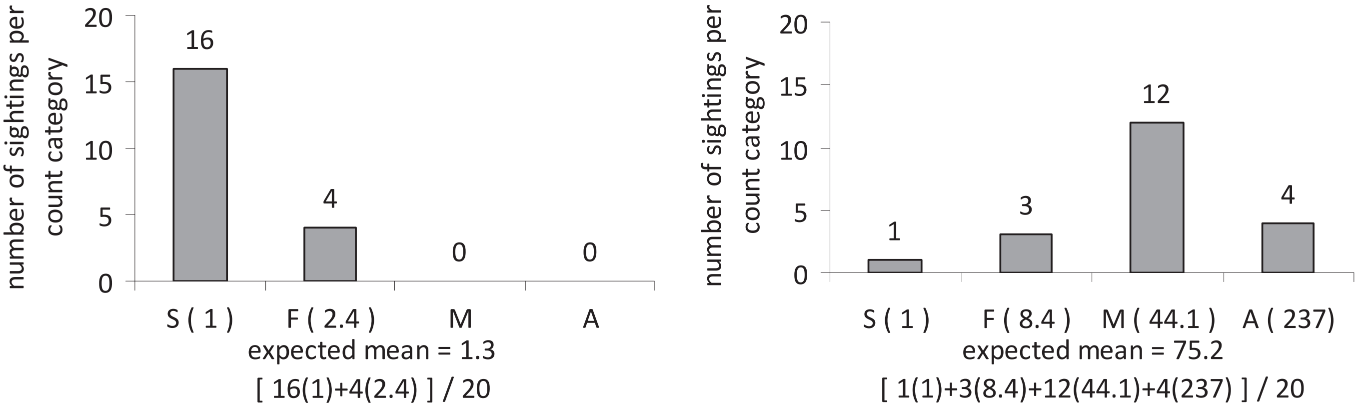

Figure 4: Distributions with a low expected mean (fig. 4a) and high expected mean (fig. 4b). Note that the expected value for "Few", *AvgF*, is 2.4 in the first distribution due to the preponderance of "Single" counts, and 8.4 in the second distribution due to the preponderance of "Many" counts.

compared using the small sample Akaike Information Criterion (AICc; Burnham and Anderson 2002), given by:

$$AICc = -n \cdot \ln(RSS/n) + 2k + \frac{2k(k+1)}{n-k-1} \qquad (9)$$

k = number of estimable parameters in the model
n = number of samples

For least squares regression, *k* is the number of parameters being fitted in the model, plus one. The added one represents RSS, the target parameter to minimize (Burnham and Anderson 2002). The sample size is $n = 36$, the number of fish species that each model was applied to and checked against. The model with the minimum AICc score has the best fit of all the models considered. Where AICc scores are very close (within 2 points), AIC relative weights can be applied.

Confidence Intervals

Confidence intervals around the expected numeric mean were calculated and combined to capture two separate sources of uncertainty: (1) SFMA translational error, and (2) observational variance. Both confidence intervals narrowed with increasing number of dive surveys.

Translational Imprecision Confidence Interval: The data upon which the estimate coefficients are based are not truly statistically independent of the data used to calculate the observed mean fish seen per survey, and therefore the regression method described above may underestimate the coefficient of variation. Furthermore, the associated coefficient of variation and confidence interval reflects an amalgamation of 36 species with behaviors that range from solitary to schooling. A bootstrap procedure was used to address issues of statistical independence, as well as determine estimate error as a function of nonzero sightings. This bootstrap procedure also avoids the issues associated with the alternative of deducing confidence intervals from log-normal distribution properties (Singh et al. 1997; Zhou and Gao 1997; Parkin et al. 1990).

The nonzero count data for the eleven most frequently seen species were broken into small segments. Fifty different segments, each *n* sightings long, were taken sequentially from a given species' string of nonzero sightings. For each segment, the prediction method (model 3) was applied and the resulting expected mean was compared to the observed mean. Each segment is so small that for all practical purposes it is statistically independent of the larger data set from which the optimal estimate coefficients were determined. This process was repeated fifty times, and the standard error was determined for this species for a given *n* sightings. The process was repeated for $n = 3$ to 150 sightings, so that a given confidence interval could be plotted against the number of nonzero sightings for each species.

To use a specific example, to determine the uncertainty associated with $n = 4$ surveys, the first four counts for 258 blue rockfish nonzero sightings were 35, 4, 6, 255. This was converted to order of magnitude counts: M, F, F, A and represents the first $n = 4$ segment of fifty segments. This segment's SFMA count is: $S = 0$, $F = 2$, $M = 1$, $A = 1$. The observed mean for the first segment was calculated (75.0) and Model 3 was applied to calculate expected mean (81.3), as well as the expected/observed ratio (81.3/75.0 = 1.084). This process was repeated for the next 49 segments of $n = 4$ length (using 200 of the 258 blue rockfish sightings), and the standard error was calculated from the fifty ratios of expected to observed means. For larger *n* values where $50n$ exceeds the total number of nonzero sightings for a species, the segments wrapped around to the beginning of the string.

The standard errors for the eleven species clustered into two main groups, "typical" species with less than

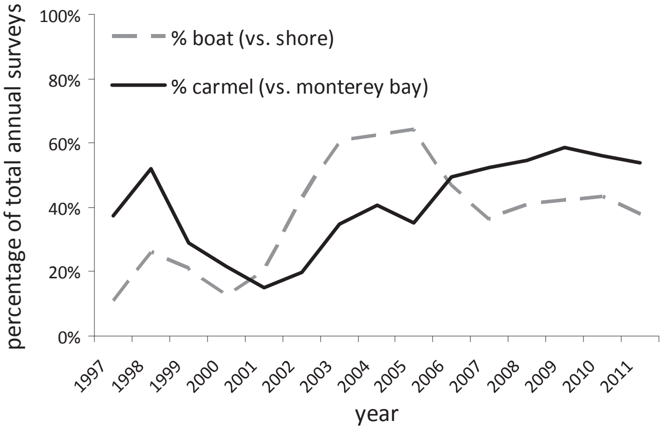

Figure 5: Year-to-year variation in dive platform (boat vs. shore) and geographic location (Carmel Bay versus South Monterey Bay) for fifteen years of REEF surveys around the Monterey Peninsula. Percentages shown are the fraction of boat dives per year (grey dashed line) and fraction of Carmel Bay dives per year (black solid line), as a fraction of total dives per year.

10% of dives reporting Abundant counts, and "abundant" or "schooling" species with a higher proportion of Abundant counts. The "abundant" species counts had less precision, both because they do not fit a log-normal distribution as well (fig. 3), and because they are not bounded by a "Super-abundant" category (over 1,000 individuals). The seven "typical" species were lingcod (*Ophiodon elongatus*), kelp greenling (*Hexagrammos decagrammus*), painted greenling (*Oxylebius pictus*), striped seaperch (*Embiotoca lateralis*), black seaperch (*Embiotoca jacksoni*), pile seaperch (*Damalichthys vacca*), and kelp rockfish (*Sebastes atrovirens*). All "typical" species reported less than 2% Abundant counts. The four "abundant" species were blue rockfish (*Sebastes mystinus*), senorita (*Oxyjulis californica*), tubesnout (*Aulorhychus flavidus*), and juvenile/young-of-year (YOY) rockfish (*Sebastes* spp). For the "abundant" species, the proportion of Abundant counts ranged from 21% to 41%.

Observational Variance Confidence Intervals: Three sources of variance in observational consistency were quantified:

1. *Variability from different divers surveying the same dive site with uneven fish distribution.* At the same dive site under the same conditions, divers will observe different numbers of fish. Part of this is due to the uneven distribution of fish across the site, and part of this is due to variation between divers in survey preferences and diving styles (e.g. path taken, swimming speed, preferred habitats). To quantify this aspect of observational variability, a study was conducted in May 2012 during a REEF Field Survey in the Monterey area where 18 experienced divers made exact counts of three common fish species (kelp greenling, striped seaperch, and blue rockfish) at two boat dive sites in Carmel Bay on one day (Outer Butterfly House and Dali's Wall), while conducting a standard REEF survey for all other species. Six coefficients of variation were calculated (2 sites x 3 species), weighted by relative proportion of nonzero sightings, and pooled by SRSS (described below in "orthogonal combination of variability").
2. *Variability from the same diver surveying the same site at different times of the year under differing conditions.* To quantify the variability from the same diver surveying the same site at different times of the year under differing conditions, the exact count data were mined for years when the same site was surveyed multiple times. Six dives in 2004 and 7 dives in 2009 were conducted at the same site in Carmel Bay (Butterfly House shore dive). Coefficients of variation were calculated for the same three common fish species as used in the variability in divers analysis. This calculated coefficient of variation was multiplied by a reduction factor to remove the effects of taking different routes across the site (quantified as described previously), and from variation caused by underlying year-to-year population trends.
3. *Variability from differing mix of dive sites.* The mix of dive sites in an area surveyed by REEF divers varies from year to year. This variability can be controlled for in three ways. First, the sites included for study can be pared down to a consistent year-to-year data set. Second, the mix of dive sites can instead be normalized for one or more characteristics to take advantage of the data for most or all sites. Third, rather than pruning or normalizing the data, the variability can also be recognized by increasing the confidence interval. This confidence interval can be estimated by taking a subset of normalized data and comparing it to the raw data. For the Monterey Peninsula area, the largest proportional variation among years was in dive platform (boat dives vs. shore dives) and geographic area (Carmel Bay vs. south Monterey Bay). For the 15-year study period, the fraction of boat dives varied among years from 11% to 64%, and the fraction of dives occurring in Carmel Bay ranged from 15% to 59% (fig. 5). For a given species, the mean number of fish sighted per year were adjusted or "normalized" to remove the variation caused by differing proportions of boat dives or Carmel Bay dives from year to year. The proportion of boat dives were normalized around the fifteen year average of 44%, while the proportion of Carmel Bay dives were normalized around the fifteen year average of 42%. The coefficient of variation between the unaltered raw data and the normalized data were calculated for three commonly seen species. The pooled three-species COV was used to increase the confidence interval to account for dive site variation.

Orthogonal Combination of Variability: The three separate sources of observational variability described above are independent of each other. Likewise, conversional imprecision and observational variability are independent of each other. Such independent sources of variability are mathematically orthogonal to each other, and are therefore combined or "pooled" by the square root of the sum of the squares (SRSS) method (Taylor 1997; Hogan 2006; NASA 2010), weighted by the relative number of observations where appropriate. The Coefficient of Variation (COV), defined in this case as standard error divided by the mean, for observational variability was then calculated as $COV/\sqrt{n-1}$ where n is the number of nonzero sightings.

RESULTS

Optimized Parameters for the Three Models

Based on minimizing the sum of Δ^2 for all 36 fish species, the best parametric fit was found for each of the three models. The optimized parameters, and resulting standard deviation and coefficient of variation of the expected-to-observed mean ratios, were as follows:

Model 1:

$$ExpectedMean_1 = 5.73^{(DEN-1)^{1.28}} \tag{10}$$

$$Std.Dev.(ExpMean_1/ObsMean) = 0.369$$

$$COV\big(SD/avg(ExpMean_1/ObsMean)\big) = 0.282$$

Model 2:

$$ExpectedMean_2 = \frac{S + 2.80F + 24.5M + 300A}{S + F + M + A}$$

$$Std.Dev.(ExpMean_2/ObsMean) = 0.105 \tag{11}$$

$$COV\big(SD/avg(ExpMean_2/ObsMean)\big) = 0.104$$

Model 3:

$$AvgF = \frac{2S + 4.16F + 10M}{S + F + M}$$

$$AvgM = \frac{11F + 33.8M + 100A}{F + M + A} \tag{12}$$

$$AvgA = \frac{200M + 348A}{M + A}$$

$$ExpectedMean_3 = \frac{S + F \cdot AvgF + M \cdot AvgM + A \cdot AvgA}{S + F + M + A}$$

$$Std.Dev.(ExpMean_3/ObsMean) = 0.085$$

$$COV\big(SD/avg(ExpMean_3/ObsMean)\big) = 0.085$$

Each of the three models, using their respective best-fit parameters, define expected/observed mean ratios for each of the 36 species, as well as summary statistics for all 36 species together (table 2).

AICc Model Comparison

The models based on disaggregated SFMA bin data (models 2 and 3) had substantially better AICc scores than the model based on an aggregate log-density index derived from the same SFMA data (model 1) (table 3). Of the two disaggregated bin data models, Model 3 scored better than Model 2 by a significant margin, with an associated AICc relative weight (probability) of 99.6% (table 3) for the models considered.

Translational Imprecision Confidence Interval

As the best model determined from AICc, Model 3 was examined in more detail. Plotting the expected mean against observed numeric mean for each of the 36 species resulted in almost a straight 1:1 line, giving a qualitative sense of the preciseness of this data conversion (figs. 6a and 6b).

From the bootstrap method, the translational confidence interval can be described as a function of n, the number of nonzero sightings, as follows:

$$\textit{Species Category Confidence Interval}_{90\%\text{ or }95\%} = a + \frac{b}{(n-1)^c} \tag{13}$$

The coefficients a, b, and c are defined according to species category (typical vs. abundant) and confidence interval (90% vs. 95% interval, that is, 5% vs. 2.5% low and high tail exclusions). Table 4 lists the coefficients used in Equation 13.

The predicted 90% confidence intervals for typical species were $^{x}/41\%$ for 10 surveys, $^{x}/29\%$ for 20 surveys, $^{x}/25\%$ for 30 surveys, $^{x}/20\%$ for 50 surveys, and $^{x}/17\%$ for 80 surveys.

At very high numbers of nonzero sightings, where the estimate error asymptotically approaches a minimum value (fig. 7), the error for the typical species category was greater than that for abundant species. Therefore, at large n sightings (more than about 130 nonzero surveys), the confidence interval for abundant species should be the maximum of either the typical or abundant species confidence interval equations.

Observational Variability Confidence Interval

Variability Between Divers: For the May 2012 study with 18 divers making exact counts of three species at two dive sites on a single day, the pooled coefficient of variation was found to be 1.11 (two dives x three species, weighted by relative number of nonzero sightings). This

TABLE 2
Comparison of models for 36 fish species based on 292 exact-count surveys.

			Expected Mean per Model			Expected/Observed Ratio per Model		
	Species Name	Observed Mean	1	2	3	1	2	3
1	Striped Seaperch	17.1	37.1	17.6	20.0	1.78	1.03	1.17
2	Blue Rockfish	114.6	123.3	117.8	119.5	0.83	1.02	1.04
3	Kelp Rockfish	12.7	21.6	11.9	12.4	1.44	0.94	0.98
4	Kelp Greenling	3.3	5.6	2.7	3.2	1.54	0.82	0.96
5	Pile Perch	8.0	10.7	8.0	7.5	1.18	1.00	0.94
6	Painted Greenling	6.3	8.8	5.7	5.6	1.25	0.91	0.89
7	Black Perch	4.6	7.4	4.7	4.6	1.45	1.02	1.01
8	Black & Yellow Rockfish	3.4	5.8	3.0	3.4	1.54	0.89	0.99
9	Blackeye Goby	6.6	9.1	7.4	6.8	1.22	1.11	1.02
10	Black Rockfish	5.6	7.2	5.1	4.8	1.16	0.91	0.87
11	Senorita	67.1	66.6	77.0	76.3	0.79	1.14	1.14
12	Lingcod	2.0	3.3	2.0	2.2	1.54	0.97	1.07
13	YOY Rockfish	125.7	81.0	106.7	107.5	0.51	0.84	0.86
14	Tubesnout	143.4	76.9	118.3	121.4	0.42	0.82	0.85
15	Olive Rockfish	4.1	5.9	4.3	4.1	1.32	1.06	1.00
16	Gopher Rockfish	3.2	4.4	2.7	2.8	1.29	0.84	0.89
17	Cabezon	1.6	2.1	1.6	1.6	1.27	1.00	0.99
18	Kelp Perch	10.3	8.3	12.1	10.6	0.72	1.17	1.03
19	Rubberlip Surfperch	5.6	7.5	5.7	5.3	1.22	1.02	0.96
20	Reef Surfperch	15.8	27.1	14.4	15.9	1.43	0.91	1.01
21	Rainbow Surfperch	10.4	10.1	8.9	8.8	0.85	0.86	0.85
22	Copper Rockfish	2.5	3.4	2.8	2.6	1.27	1.10	1.02
23	Snubnose Sculpin	1.7	2.4	1.7	1.7	1.36	1.01	1.02
24	Speckled Sanddab	4.0	6.8	4.6	4.4	1.54	1.15	1.12
25	Yellowtail Rockfish	3.4	4.8	4.0	3.6	1.31	1.18	1.08
26	Gibbonsia Kelpfish	1.2	1.3	1.2	1.2	1.07	1.00	0.94
27	Treefish	1.4	1.8	1.5	1.4	1.23	1.04	0.99
28	Monkeyface Prickleback Eel	1.4	1.8	1.5	1.4	1.31	1.10	1.06
29	Vermillion Rockfish	1.1	1.3	1.2	1.1	1.14	1.08	1.02
30	Opaleye	2.8	4.3	3.3	3.1	1.43	1.19	1.12
31	California Sheephead	2.2	3.8	2.1	2.4	1.62	0.94	1.08
32	Blacksmith	68.0	18.4	77.0	79.6	0.23	1.13	1.17
33	Scalyhead Sculpin	1.4	1.8	1.5	1.4	1.25	1.04	1.01
34	Grass Rockfish	1.3	1.5	1.3	1.3	1.13	1.02	0.96
35	Kelp/Calico Bass	1.3	1.6	1.4	1.3	1.19	1.05	0.99
36	Coralline Sculpin	1.3	1.6	1.4	1.3	1.26	1.10	1.05
	Standard Deviation of (Expected Mean / Observed Mean)					0.338	0.105	0.085
	Coefficient of Variation [SD / mean(Expected Mean / Observed Mean)]					0.282	0.104	0.085
	RSS/n					0.1746	0.0110	0.0071

Note: RSS = sum of ln (Expected Mean / Observed Mean) squared

TABLE 3
Akaike Information Criterion (AICc) comparison of Models. The model with the lowest AICc value indicates the best balance between under- and over-fitting, with best relative likelihood (probability).

Model	RSS/n	k	AICc	AIC Weight, w_i (probability)
Model 1	0.1746	3	–54.3	0.000
Model 2	0.0110	4	–153.1	0.001
Model 3*	0.0069	9*	–154.2	0.003
Model 3	**0.0071**	**5**	**–161.1**	**0.996**

* A variation of Model 3 where the *a priori* constants 2, 10, 11, and 100 are allowed to vary to more closely fit observed data. Per AICc, this approach, which increases the number of fitting parameters to minimize RSS, overfits the data.

captures the variability from divers with different surveying habits, preferences and swimming speeds, taking different routes across the same dive site, and counting fish that move about and are found in patches across the site.

Variability From Differing Conditions Over a Year: Looking at three common fish species, the pooled coefficient of variation for differing conditions over the year was found to be 0.79. However, part of this variation was not random, but instead reflected real population trends. Another part of this variation, surveying different areas at the same site, overlaps and repeats the variation measured in the May 2012 study. Assuming that 40% of this variation is due to differing conditions over the year at the same site, this coefficient of unique variation is estimated to be 0.32.

Variability From a Changing Mix of Dive Sites: Year-to-year fluctuations in fish populations, normalized for Monterey Bay versus Carmel Bay dives (fig. 5), and shore versus boat dives (fig. 5), were compared against the raw data (fig. 8). Calculating the coefficient of variation for each of the three species, and pooling appropri-

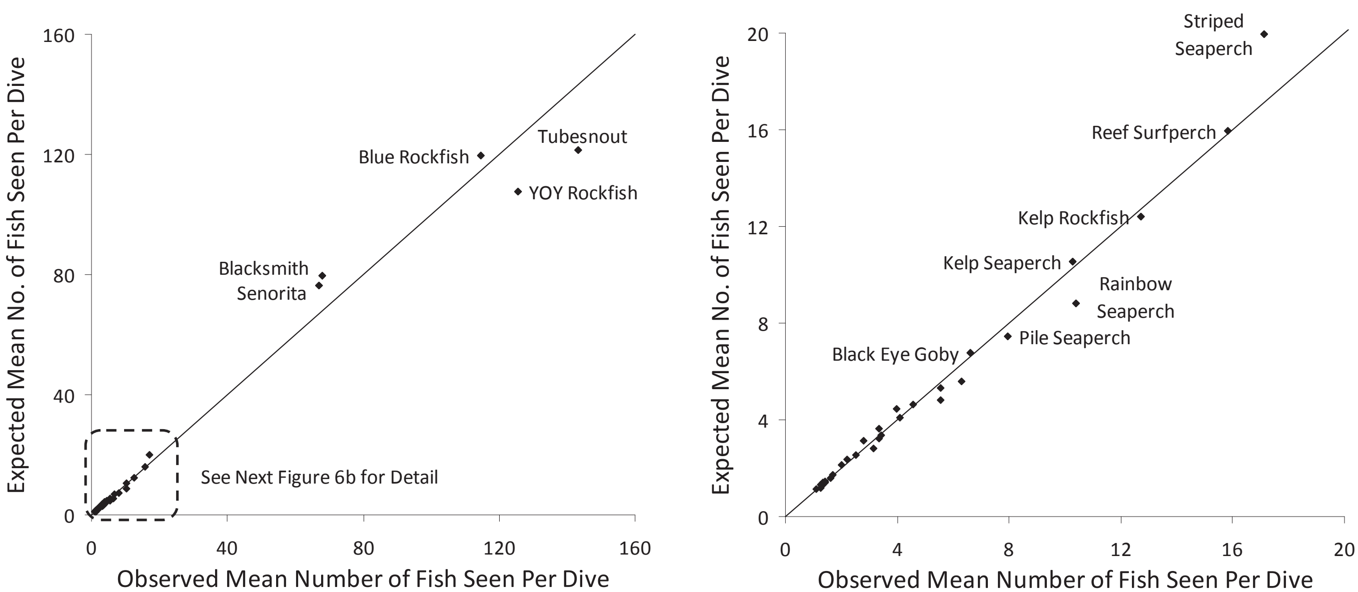

Figure 6: Model 3's expected versus observed mean number of fish sighted per dive for various species (each data point represents an observed fish species). Data from a perfect predictive model would fall on a 1:1 line. Figure 6a (left graph) shows the full range of species' predicted and observed counts (0–160). Figure 6b (right graph) shows that most species cluster in the region of 20 or fewer average sightings per dive.

TABLE 4
Coefficients for Calculating Confidence Interval, $C.I. = a+b/(n-1)^c$ for Model 3, where n = number of nonzero fish sightings for a given species, area, and time period.

Species Category:	Typical Species		Abundant Species	
Proportion of Abundant Counts:	Less than 10%		More than 10%	
	TRANSLATIONAL ERROR			
Confidence Interval:	90%	95%	90%	95%
Coefficients:				
a	0.05	0.06	–1.80	–2.16
b	1.07	1.28	3.53	4.22
c	0.50	0.50	0.12	0.12
	TRANSLATIONAL + OBSERVATIONAL ERROR			
Confidence Interval:	90%	95%	90%	95%
Coefficients:				
a	0.02	0.03	–0.26	–0.27
b	2.03	2.42	3.20	3.88
c	0.48	0.48	0.38	0.39

Translational Error = Uncertainty from converting SFMA $\log_{10}$-bin count data to expected arithmetic mean.

Observational Error = Uncertainty from stochastic distribution of fish across a dive site, and variability between divers, diving conditions, and dive sites.

ately, leads to a coefficient of variation of 0.15.

Observational Variability Pooled Across Sources: The pooled coefficient of variation for the three sources of observational error is therefore estimated to be

$$\sqrt{1.11^2 + 0.32^2 + 0.15^2} = 1.16.$$

Observational standard error (normalized to one) as a function of n surveys is $\pm 1.16/\sqrt{n-1}$. Because this is a log-normal distribution, it is more accurate to express the standard deviation relative to the mean as the mean multiplied or divided by $\upsilon + \sqrt{1+\upsilon^2}$ where $\upsilon = 1.16/\sqrt{n-1}$

Combined Translational and Observational Confidence Intervals

The observational error can be combined with the SFMA translational error using SRSS to express the combined error as a function of the number of nonzero sightings in a year for two fish species categories, typical and abundant (fig. 9). The same combined error is also expressed as coefficients of Equation 13 in Table 4. For typical fish species, the 90% confidence interval drops below $^{\times}/30\%$ at 60 sightings per year, and the 95% confidence interval drops below $^{\times}/30\%$ at 90 sightings per year.

DISCUSSION

Models for Estimating Abundances: The methods described in this paper convert order-of-magnitude counts to an aggregate arithmetic mean with a reasonably small confidence interval, given a sufficient number of surveys. Parameters for each model were optimized for best fit to 36 fish species common to the Monterey Peninsula in California. The distributions of exact counts from 292 surveys for these 36 fish species were essentially log-normal. Because these distributions were generated by roving diver "random walks" across patchy distributions of fish on nearshore reefs, the optimized estimation parameters, based on underlying stochastic distributions, are likely to also apply to other species and geographic regions.

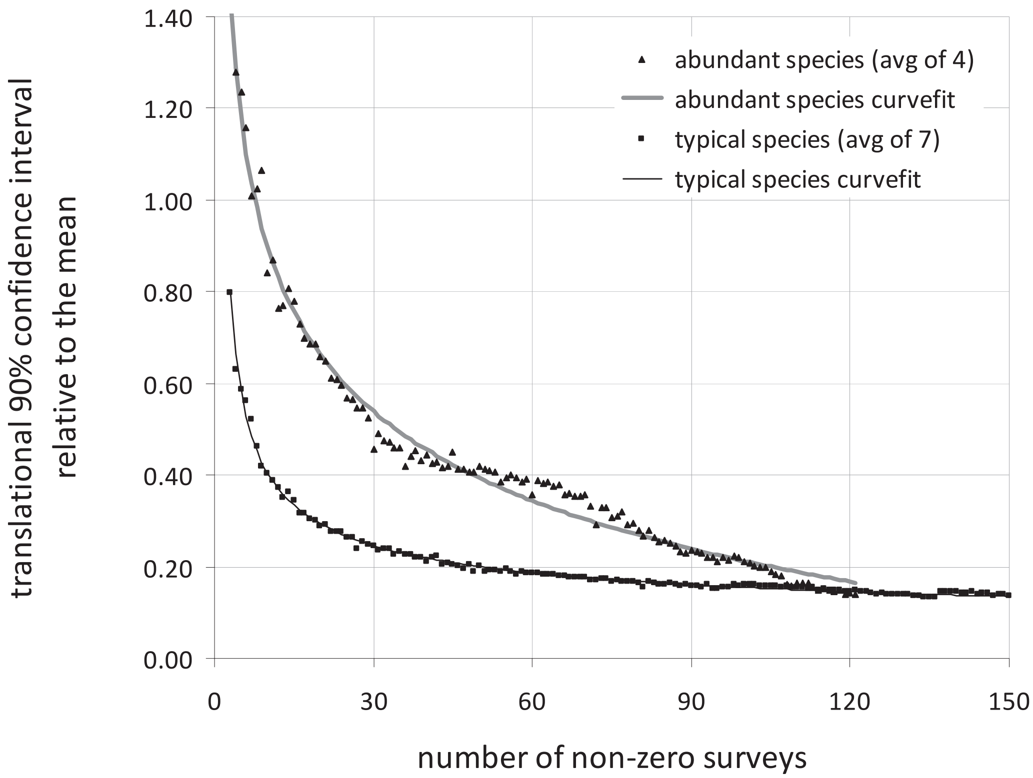

Figure 7: Translational 90% confidence interval relative to the mean (vertical axis), as a function of number of non-zero surveys (horizontal axis). Translational error is that associated with converting SFMA data to expected arithmetic means. A confidence interval value of 30% indicates the upper end of the interval is mean times 1.30, and the lower end is mean divided by 1.30. "Typical" species are those where Abundant sightings comprise less than 10% of all nonzero sightings. "Typical" species error is based on the average of 7 typical species. "Abundant" species are those where Abundant sightings comprise more than 10% of all nonzero sightings. "Abundant" species error is based on the average of 4 abundant species.

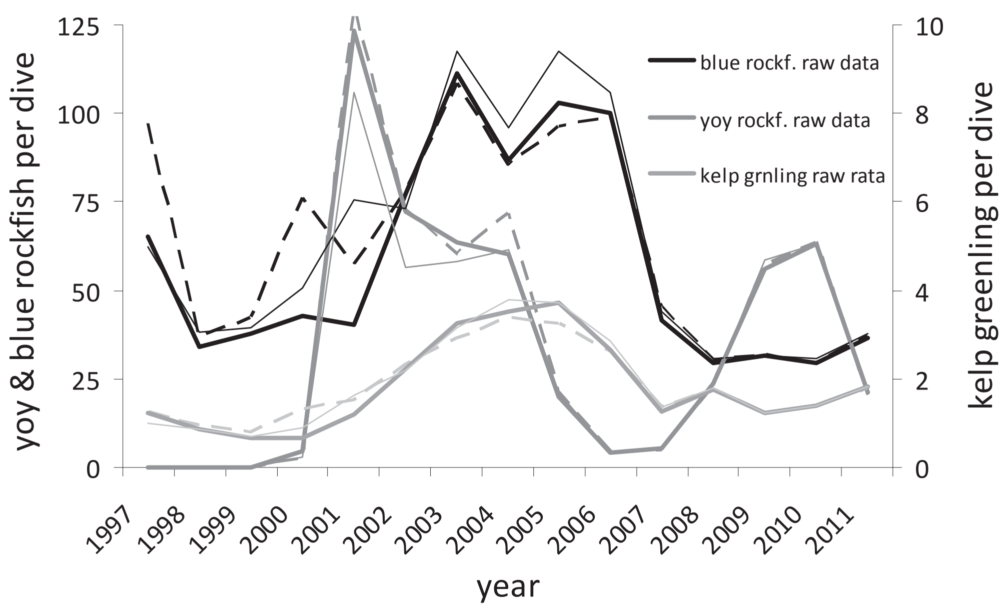

Figure 8: Fish population estimates for three species (YOY rockfish < 5 cm, blue rockfish > 5 cm, and kelp greenling) are plotted over 15 years, based on REEF surveys around the Monterey Peninsula, California. For purposes of comparison, raw data estimates are plotted alongside data normalized for consistent year-to-year boat/shore and Carmel/Monterey ratios. Solid lines are raw data, dashed lines are normalized for boat versus shore dive sites, and thin lines are normalized for Carmel versus Monterey bay dive sites. Refer back to Figure 5 for the year-to-year variations in mix of dive sites that create the differences between raw and normalized data.

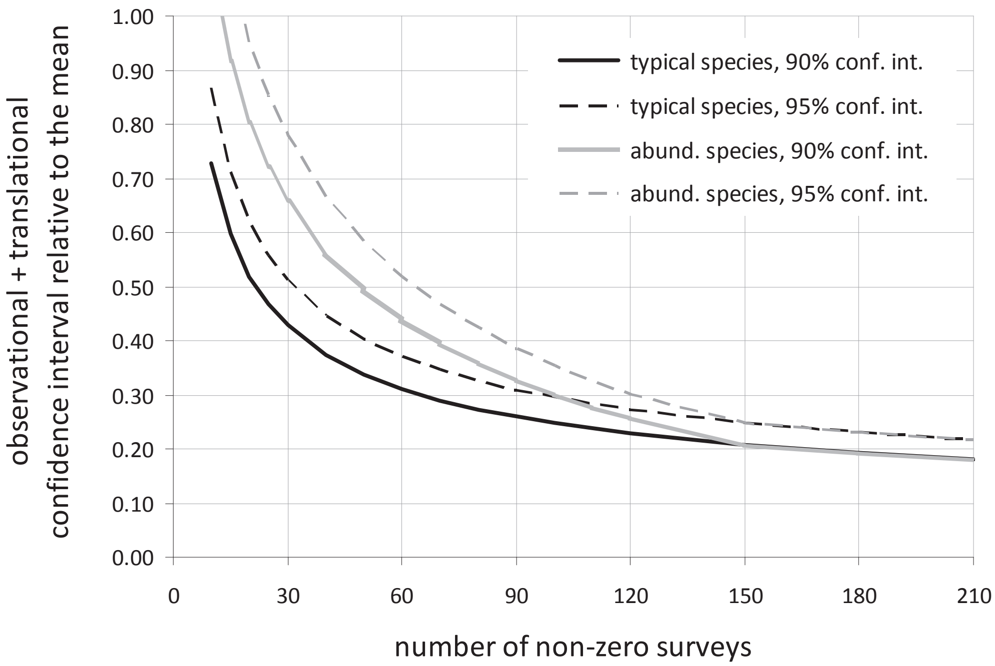

Figure 9: The combined observational and translational confidence interval (vertical axis) as a function of the number of nonzero dive surveys (horizontal axis. Translational error is that associated with converting SFMA data to expected arithmetic means, while observational error is from uneven fish distributions and variations between divers, dive conditions and dive sites. 90% and 95% confidence intervals corresponds to 5% and 2.5% high and low tails, respectively. A confidence interval value of 30% indicates the upper end of the interval is mean times 1.30, and the lower end is mean divided by 1.30. "Typical" species are those where Abundant sightings comprise less than 10% of all nonzero sightings. "Typical" species error is based on the average of 7 typical species. "Abundant" species are those where Abundant sightings comprise more than 10% of all nonzero sightings. "Abundant" species error is based on the average of 4 abundant species.

Both the coefficient of variation of the predicted/observed ratios and AICc scoring suggests that Model 3 is the most accurate in converting SFMA data to numeric means. Model 3 appears to be the most effective because it (a) considers the relative proportion of adjacent count bins, and (b) reflects the underlying lognormal distribution of fish sightings. In fact, the exact count data indicate that the f_f and m_m coefficients are close to (within 5% of) the mean of the log-transformed bounds for their respective ranges. That is:

$$f_f = 4.16 \approx e^{(\ln(1.5) + \ln(10.5))/2} = 3.97 \quad (14)$$

$$m_m = 33.8 \approx e^{(\ln(10.5) + \ln(100.5))/2} = 32.5 \quad (15)$$

The values for these coefficients make intuitive sense because the sightings in the integer sub-bins (e.g., 2, 3…9, 10) within each range (e.g., Few) decline exponentially. Because the Abundant category is the highest order of magnitude counted, the a_m and a_a coefficients do not follow the same pattern.

Per the AICc selection criterion, Model 3 prevails over Model 2 given the assumed values for k, the number of parameters, of 5 and 4, respectively. Some may argue that the constants 2, 10, 11, and 100 found in Estimate 3 are actually AIC parameters and should be accounted for in k when calculating AICc scores. The authors considered these *a priori* constant properties based on the definition of Few (2–10) and Many (11–100). In fact, a slightly smaller apparent confidence interval can be attained if these constants are allowed to vary as fitting parameters. However, AICc analysis revealed that allowing these constants to vary ($k = 9$) is inappropriate overfitting.

Even though Model 2 does not prevail according to AICc scoring, it may nevertheless be attractive to some researchers because of its simple formation and reasonable accuracy, with a sacrifice of only about 22% in increased confidence interval compared to Model 3.

Model 1, with a confidence interval roughly three times larger than Models 2 and 3, can still be useful in making rough estimates and discerning global trends, when the only data available are the sighting frequency (*SF*) and log-density index (*DEN*) information available to the public on the www.REEF.org Web site. Disaggregated SFMA data, upon which Models 2 and 3 rely, are available to researchers through special request.

Some researchers may propose ordered logit regression as an alternative estimation method. A preliminary investigation suggests that this approach will achieve AICc scores better than Model 1, but worse than Models 2 and 3. This might be expected, because the method is

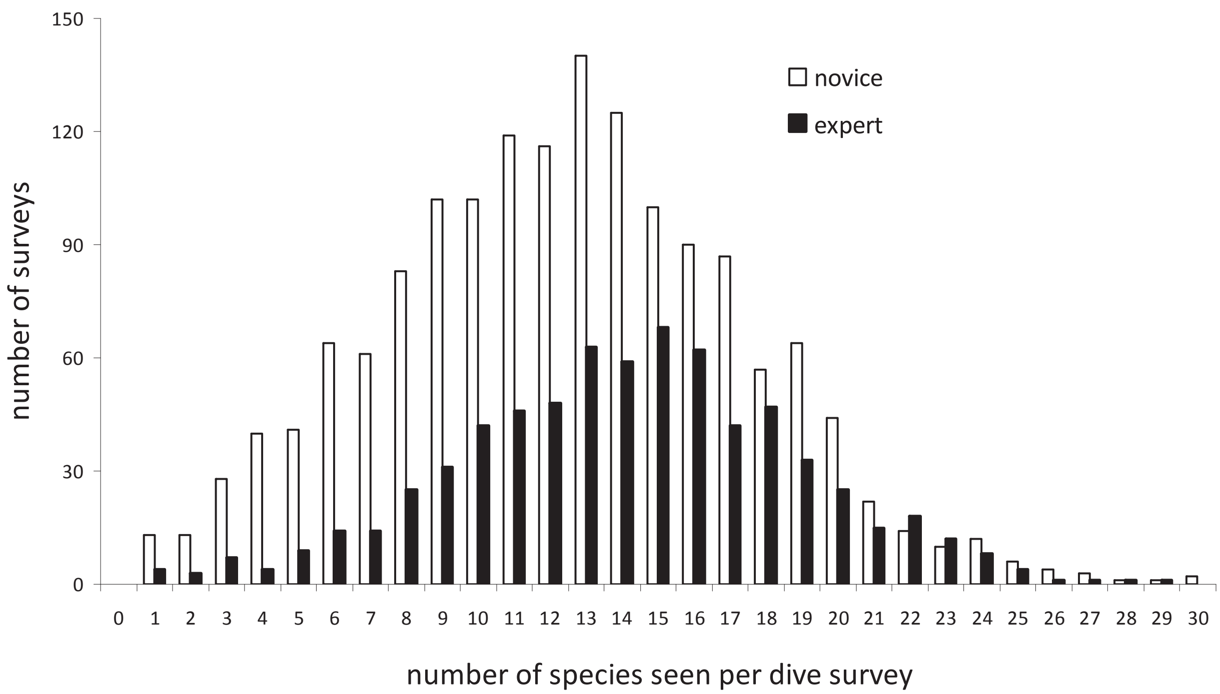

Figure 10: Histogram of number of surveys (vertical axis) as a function of number of fish species seen per dive (horizontal axis), for novice surveyors (white bars) and expert surveyors (black bars). The mean number of fish species seen on a dive by novices is 12.4, compared to 14.1 species for experts.

designed for analyzing qualitative ranges such as opinions that range from "strongly disagree" to "strongly agree." Because the logit function is based on the premise of evenly distributed categories, other ordinal regression (cumulative probability) approaches may prove more fruitful, such as negative log-log, based on the more accurate assumption that lower categories are more probable (Norusis 2011).

Dive Site Mix Normalizing Versus Increased Confidence Interval: The reported observational error to take into account a varying mix of dive sites (COV = 0.15) is specific to the Monterey Peninsula data, to the dive site characteristics considered, and to the three species considered. As such, it only represents what may be a "typical" value for this kind of variability in REEF surveys. It is always preferable to instead control the dive site mix, by either culling dive site data to maintain a consistent mix, or to normalize a varying mix of dive sites. When the dive site mix is not controlled or normalized, it is preferable to use COV = 0.15 rather than ignore this contribution to variability altogether.

Bias Towards Undercounting. The REEF RDT survey method may have a consistent tendency to undercount fish populations, leading to conservative estimates of absolute fish densities if this tendency is not accounted for in analyses. However, because this bias is consistent over time, underlying population dynamics should be discernable even if the bias is not corrected. The undercount bias is due to several factors:

1. Strict instructions to surveyors to not count any fish seen whose species cannot be positively identified (Pattengil-Semmens and Semmens 2003a). Novice surveyors therefore undercount less readily identified species (fig. 10). However, previous research suggests that novices count the number of individuals of species that they can identify as accurately as experts (Pattengill-Semmens and Semmens 1998).
2. Tendency to overlook cryptic bottom dwelling species as well as fish sheltering in concealed unobservable crevices.
3. Tendency to not notice pelagic fish swimming overhead or those quickly swimming in and out of view.
4. Inability to see sufficient distance in low visibility conditions.
5. Inability to thoroughly examine the bottom in high surge conditions.
6. Tendency of divers to not count the same fish twice on the return leg of their route (affects density estimates that are based on estimated length of dive route).
7. Tendency of divers to undercount schools of fish. Harrison Stubbs (pers. comm. June 2012) has observed that most people consistently underestimate the number of fish in a still photo of schooling fish,. Even greater bias may occur in situ where a school is constantly moving. This could explain the truncated log-normal distribution at higher fish counts shown in Figure 1b, as well as the higher deviations at higher counts shown in Figure 3.

Sightings Per Dive as an Estimate of Population Density: An underlying premise of this paper is that the mean number of fish seen per dive can be used as a proxy for nearshore fish species relative abundance. This is reasonable, given that dives have an average duration and route length. While this creates internally consistent data for comparisons between areas and between time periods, it is also useful to make estimates of absolute density to compare to other data sets. Wolfe and Pattengill-Semmens 2013 evaluated REEF data collected between 1997–2011 for the Monterey Peninsula area, and compared year-to-year fluctuations with exact-count transect data collected by the Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO). The two data sets correlated well. Furthermore, REEF data could be calibrated with the transect data. For most fish species, a rough estimate of density can be determined from REEF count data by dividing the average number of fish seen per dive by 300 m^2. Because a REEF diver typically surveys a larger area than this on a given dive (approximately 50 minute average dive x 5 m/min average swim speed x 2 m minimum wide swath = approximately 500 m^2), the comparison of PISCO and REEF data provided indirect evidence that the RDT method has some bias toward undercounting compared to transect surveys.

Conclusion: The methods described in this paper convert order-of-magnitude counts to an aggregate arithmetic mean with a reasonably small confidence interval if the number of surveys is sufficiently large. These methods hold particular potential for application to REEF fish surveys. Researchers using RDT data have typically used Abundance Score indices, which is sighting frequency (*SF*) multiplied by the log density index (*DEN*), sometimes calibrated to transect surveys (Green et al. 2012; Pattengill-Semmens and Semmens 2003b). The model proposed in this paper offers two significant advantages: (a) fish seen per dive can be reported in arithmetic units that are easier to compare to other data sets, and (b) confidence intervals are much tighter. For instance, Green et al. 2012, plotting fish seen per dive on a log scale, reported a 95% confidence interval for lionfish in 2008 based on 21 surveys of about $^{x}/3.55$, with similarly broad intervals in other years. Model 1 in this paper would have similar SFMA translational confidence intervals. However, using Model 3, the 95% confidence interval would be $^{x}/1.35$ considering only SFMA translation error, and $^{x}/1.61$ if both translational and observational error were considered—a dramatic tightening of confidence interval in either case.

REEF data for the Monterey Peninsula area of the California coast were used as an example, with results described in a companion paper (Wolfe and Pattengill-Semmens 2013). Both the conversion error and observation error appeared to be reasonably small. For the Monterey data, with annual number of surveys ranging from 50 to 353, it appeared that 90% confidence intervals (5% high and low tails) of combined SFMA conversion and observation error were typically less than $^{x}/30\%$ relative to the mean. For fish population estimates, where confidence intervals are often very broad, this demonstrates that fish population estimates with reasonable error bars can be attained at remarkably low cost with citizen science volunteer efforts supported by a small professional staff.

The authors believe that the methodology of converting SFMA data to numeric means described in this paper can be applied to other coastal ocean areas where large numbers of REEF surveys have been conducted. In addition to the Monterey area of central California, REEF maintains a long-term database of surveys for many areas of the coastal United States and Caribbean, including the Pacific Northwest, Southern California, Hawaii, the South Pacific, and the tropical western Atlantic. The estimation method described in this paper should also prove useful in analyzing data from other order-of-magnitude population survey efforts.

ACKNOWLEDGEMENTS

Paul Humann and Ned DeLoach founded REEF in 1990, and Lad Akins served as its executive director for many years—fostering the efforts of thousands of recreational divers to conduct fish and invertebrate surveys. Brice Semmens was instrumental in developing REEF's Roving Diver Technique and has served as a valuable scientific advisor. Lynn Waterhouse, Tom Laidig, Robert VanBuskirk, Nathan Hunt, and Katie Coughlin provided insightful statistical comments. Hundreds of volunteer divers have contributed to the REEF database for Monterey Peninsula. We especially thank the 18 members of the Advanced Assessment Team who assisted us with the Variability Between Divers study in May 2012.

LITERATURE CITED

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodal inference, 2nd Ed. Springer. 488 pp.

Engen, S., and R. Lande. 1996. Population dynamic models generating the lognormal species abundance distribution. Math. Biosci. Vol. 132:169–183.

Green S. J., J. L. Akins, A. Maljković, and I. M. Côté. 2012. Invasive lionfish drive Atlantic coral reef fish declines. PLoS ONE 7(3): e32596. doi:10.1371/journal.pone.0032596.

Hogan, R. 2006. How to Combine Errors. Teaching Resource, University of Reading, UK. (www.met.rdg.ac.uk/~swrhgnrj/combining_errors.pdf).

Holt, B. G., R. Rioja-Nieto, M. A. MacNeil, J. Lupton, and C. Rahbek. 2013. Comparing diversity data collected using a protocol designed for volunteers with results from a professional alternative. Methods in Ecology and Evolution, 4: 383–392.

Limpert, E., W. Stahel, and M. Abbt. 2001. Log-normal Distributions across the Sciences: Keys and Clues. Bioscience. 51(5): 341–352.

May, R. M. 1975. Patterns of species abundance and diversity. *In* Ecology and Evolution of Communities, M.L. Cody and J.M. Diamond, eds. Harvard University Press, pp.81–120.

NASA 2010. NASA Handbook: Measurement Uncertainty Analysis Principles and Methods. NASA-HDBK-8739.19-3.

Norušis, M. J. 2011. IBM SPSS Statistics 19 Advanced Statistical Procedures Companion. Prentice Hall. 464 pp.(www.norusis.com/book_ASPC_v19.php).

Parkin, T. B., S. T. Chester, and J. A. Robinson. 1990. Calculating confidence intervals for the mean of a lognormally distributed variable. Soil Sci. Soc. Am. J., 54: 321–326.

Pattengill-Semmens, C. V., and B. X. Semmens. 1998. An analysis of fish survey data generated by nonexperts in the Flower Garden Banks National Marine Sanctuary. J. Gulf Mex. Sci. 2: 196–207.

Pattengill-Semmens, C. V., and B. X. Semmens. 2003a. Conservation and management applications of the REEF volunteer fish monitoring program. Env. Monitor. Assess. Journal. 81: 43–50.

Pattengill-Semmens, C.V., and B. X. Semmens. 2003b. The status of reef fishes in the Cayman Islands (B.W.I.). Atoll Res. Bull. Vol. 496, pp. 226–247.

REEF. 2013. Reef Environmental Education Foundation (REEF) Online Database. www.REEF.org/db/reports.

Schmitt, E. F., T. D. Sluka, and K. M. Sullivan-Sealy. 2002. Evaluating the use of roving diver and transect surveys to assess the coral reef assemblages off southeastern Hispaniola. Coral Reefs. 21: 216–22.

Semmens, B. X., J. L. Ruesink, and C.V. Pattengill-Semmens. 2000. Multisite multi-species trends: a new tool for coral reef managers. Proc. 9th Int. Coral Reef Symp., October 2000. 1071–1078.

Singh, A. K., A. Singh, and M. Englehardt. 1997. The lognormal distribution in environmental applications. EPA Technology Support Center Issue (EPA/600/S-97/006).

Taylor, J. R. 1997. An introduction to error analysis: the study of uncertainties in physical measurements. 2nd Ed. University Science Books. 327 pp.

Wolfe, J. R., and C.V. Pattengill-Semmens. 2013. Fish Population Fluctuation Estimates Based on Fifteen Years of REEF Volunteer Diver Data for the Monterey Peninsula, California. Calif. Coop. Oceanic. Fish. Invest. Rep. 54: 141-154.

Zhou, X. H., and S. Gao. 1997. Confidence intervals for the log-normal mean. Statistics in Medicine. 16: 783–790.