PROSPECTIVE (A PRIORI) POWER ANALYSIS FOR DETECTING CHANGES IN DENSITY WHEN SAMPLING WITH STRIP TRANSECTS

KONSTANTIN A. KARPOV Karpov Marine Biological Research 24752 Sashandre Lane Fort Bragg, CA 95437 e-mail:kkarpov1@gmail.com

> MARY BERGEN 11584 Creek Road Ojai, CA 93023

JOHN J. GEIBEL California Department of Fish and Game 425 Central Avenue Menlo Park, CA 95025

PHILIP M. LAW California Department of Fish and Game 350 Harbor Boulevard Belmont, CA 94002

CHARLES F. VALLE California Department of Fish and Game 4665 Lampson Avenue, Suite C Los Alamitos, CA 90720

DAVID FOX Oregon Department of Fish and Wildlife 2040 Southeast Marine Science Drive Newport, OR 97365

When designing a monitoring program, it is important to determine how much sampling is needed prior to data collection. Programs with too little statistical power produce ambiguous results and public debate that cannot be resolved. However, prospective power analysis requires an estimate of sample variance. In this paper, data from strip transect surveys using remote operated vehicle (ROV) of fish on temperate subtidal rocky reefs were used to establish the relationship between density and variance needed for power analysis. The relationship was used to select the optimal sample unit (transect) size and estimate the total sampling effort needed to measure specific changes in density between two sampling study areas. In general, smaller transects were more efficient than larger transects. The smallest transects (50 m²) were most efficient, but the difference between 50-m² transects and 100-, 200-, and 400-m² transects was relatively small (11% to 28%). The largest transects (800 m²), however, required 57% more sampling area than 50-m² transects. The total sampling area needed to detect a significant difference in density increased with decreasing effect size, as expected. Also as expected, some species (e.g. copper rockfish) required more sampling effort than others (e.g. vermilion rockfish). These results demonstrate that pre-existing data may be used to establish relationships between means and variances, and to determine the optimal transect size and the amount of sampling effort needed to measure statistically significant differences in fish density between study areas.

Key Words: monitoring, power analysis, sampling design, statistical power, transect sampling

INTRODUCTION

When designing a monitoring program, it is important to determine how much sampling is needed prior to data collection. A finding of "no significant difference" in light of insufficient power produces ambiguous results that cannot be resolved due to failure of rejecting false null hypotheses (Toft and Shea 1983; Hayes 1987; Peterman 1990). Calculating power *a posteriori* does not solve the problem (Steidl et al. 1997). Questions, statistical design, and scope of sampling can be tailored to the available budget if power is considered *a priori*. Estimating statistical power of a given sampling design will clarify which questions can be answered, and can provide information for setting priorities with different levels of funding. A monitoring program with clearly formulated questions, appropriate statistical design, and sufficient statistical power will increase the probability of answering the most important questions.

However, prospective power analysis requires an estimate of sample variance. Because variance is influenced by multiple factors, including study area selection and temporal trends, it is difficult to predict accurately. It is possible, however, to put bounds on the variance. Steidl et al. (1997) suggested using pilot studies, values from similar research in other geographic areas, or a range of probable values. Gibbs et al. (1998) used values for variance from published literature in Monte Carlo simulations based on linear regression to compute sample sizes for measuring change over time in taxonomic groupings of grasses, sedges, and large mammals. Carr and Morin (2002) used values in published literature with linear regression to compute sampling effort for studies of bacterial abundance and production.

The objective of this study was to select the optimal sample unit (transect) size and estimate the total sampling effort needed to measure specific changes in density between two sampling study areas for remotely operated vehicle (ROV) surveys of fish on temperate subtidal rocky reefs. Because of logistics, ROV surveys are generally done on long track lines (Barry and Baxter 1993). The long lines are then broken into transects. However, the choice of transect size has generally been arbitrary. Herein, we use existing ROV survey data with a full range of variability to establish the relationship between density and variance needed for power analysis. The power analysis is used to select the optimal transect size and sampling effort.

METHODS

The power analyses presented here were used in designing ROV surveys of fish populations, one element of a monitoring program to evaluate marine protected areas (MPAs) in California (http://www.dfg.ca.gov/marine/channel_islands). The MPA monitoring program included quantifying habitat and density of fish in various protected and reference areas over time. A primary objective of the ROV surveys was to measure changes in density of demersal fish populations in deep water (10 to 70 m) hard bottom habitats between single or combined study areas by treatment (fished vs. MPA).

To represent a likely range of density and variance, strip transect data were collected from 12 study areas (Figure 1) spanning a large geographic area collected by two research groups using similar methods. Data were compiled from separate surveys of three areas: 1) Siletz Reef on the central Oregon coast, conducted by the Oregon Department of Fish and Wildlife (ODFW) in 2002; 2) MacKerricher State Park near Fort Bragg in northern California, conducted by the California Department of Fish and Game (DFG) in 2004; and 3) 10 study areas at the northern Channel Islands off the coast of southern California, conducted by DFG in 2006. Data from different surveys were used so that variance would include all sources of error, including survey methodology and sampling error.



Figure 1. Locations of 12 ROV study areas used in this study: Siltez Reef in Oregon, MacKerricher State Park in northern California, and ten study areas among the northern Channel Islands in southern California.

Siletz Reef was the largest study area, spanning 20 km of coastline, with both contiguous and isolated rocky reefs. The study area in MacKerricher State Park spanned 1.5 km of coastline. The 10 study areas at the northern Channel Islands (Figure 2) varied in size from 500 m to about 1 km of coastline. The average depth of sampling ranged from 22 to 47 m (Table 1). Substrate topography at 5 of the 12 study areas was classified as low (< 2 m) to medium (2 to < 3 m) relief, medium to high relief (\geq 3 m) at 6 study areas, and a mixture of all three categories at one study area. Except for one study area, the northern Channel Islands had a higher percentage of soft substrate (sand, gravel, or cobble) than study areas in northern California and Oregon; 8 of 10 study areas in the northern Channel Islands had more than 34% soft substrate.



Figure 2. Detailed locations of the 10 ROV study areas in the northern Channel Islands. Study area names and codes are listed on Table 1. State Marine Reserves (SMRs) are marine protected areas closed to all fishing. State Marine Conservation Areas (SMCAs) allow limited fishing.

Survey Design and Sampling

Survey protocols were designed to videotape long continuous strip transects of known length and width by using sonar linked to GPS tracking (Karpov et al. 2006) along preplanned target lines. In both Oregon and California, target lines across hard substrate were selected using habitat interpretation of side-scan or multibeam sonar maps. Target lines, positions of the ROV and the ship, water depth, and distance from the ROV to the substrate were displayed on navigational computer monitors. The ROV pilot maintained a forward course along the target line while the ship's captain maintained position relative to the ROV. These protocols produced transect lengths accurate to at least 2 m as tested across distances of 6 to 100 m (Karpov et al. 2006).

Three different methods were used to distribute target lines across study areas. At Siletz Reef, ODFW used a stratified random design to allocate target lines, 406 to 905 m in length, into two depth strata (5-30 m and 31-60 m) and two strata of relative topographic

Table 1.	Depth, relief, and substrate typ	pe for the 12 sampled study areas.	All study areas
a	re in fished areas unless design	nated as a State Marine Reserve (Sl	MR) or State
Ν	Marine Conservation (SMCA).	Study area codes are used in Figure	re 2.

	Depth			Relief ^a		
Stud area location, name and	(m)			(Low,	Percent	
(code)				Medium,	Hard &	Percent
	Mean	Min	Max	High)	Mixed	Soft
Central Oregon						
Siletz Reef	33	13	52	М, Н	75 ^b	25 ^b
Northern California						
MacKerricher State Park	22	14	30	L, M	86	14
Channel Islands, California						
San Miguel Island						
Harris Point SMR (SMI 1)	41	19	60	М, Н	57	43
Castle Rock (SMI 2)	44	18	65	М, Н	91	9
Santa Rosa Island						
Carrington Pt. SMR (SRI 2)	31	16	46	L, M	64	36
Rodes Reef (SRI 3)	29	18	45	L, M	72	28
Cluster Point (SRI 7)	27	14	53	М, Н	65	35
South Point SMR (SRI 8)	36	16	59	М, Н	49	51
East Point (SRI 6)	45	17	74	L, M, H	47	53
Santa Cruz Island						
Gull Island MPA (SCI 2)	47	27	66	М, Н	37	63
Anacapa Island						
Anacapa Island MPA (AI 1)	34	12	54	L, M	56	44
Anacapa Island MPA (AI 2)	43	18	72	L, M	51	49

^a Low < 2 meters, Medium 2 to < 3 m, High ≥ 3 m

^b Substrate classifications used by ODFW and CDFG differed (see Methods). ODFW values were modified to equal DFG values.

relief (high and low relief) (see Figure 1). At MacKerricher State Park, DFG systematically placed target lines in two rocky areas separated by approximately 500 m; these lines were 500 m long, separated by 100 m, and parallel to the shoreline (see Figure 1). At the northern Channel Islands, DFG randomly placed target lines 500 m in length, separated by at least 20 m, in up to four rectangular zones (Figure 2). The number of lines per zone was weighted according to the amount of expected hard substrate. Data were used only when the ROV was on the targeted track line making forward progress and the laser lights were visible on the bottom at the Siletz Reef study area (ODFW), or when the distance to the substrate was within 4 m as measured by ranging sonar at MacKerricher and Channel Island study areas (Karpov et al. 2006).

Transect length was computed from navigational data using running averages with 7 points for ODFW and 21 points for DFG data (Karpov et al. 2006). Lengths computed with 7 and 21 points are equivalent (Karpov et al. 2006). ODFW computed transect width from measured distance between lasers projected on the substrate (Wakefield and Genin 1987). DFG computed transect width from the distance to the substrate measured with ranging sonar and properties of the camera (Karpov et al. 2006). The two measurements are equivalent when the ROV is within 4 m of the substrate (Karpov et al. 2006). Transect width averaged 3 m for the 12 areas.

Fish in the video film were enumerated by discernible taxa (e.g. species, species complex, family, or unidentified) in a swath approximately equal to measured transect width in plane with paired lasers or ranging altimeter. ODFW counted fish in the lower 80% of the video screen. DFG counted fish with the aid of a transparent film overlay on the top half of the video screen monitor. Two converging guidelines on the transparency approximated the vanishing perspective of the strip transect based on the camera tilt angle relative to the forward plane of view. Only fish that were at least halfway within the transect guidelines and within 4 m of the camera were counted. Distances from camera were based on sonar range values depicted on the screen. Fish smaller than the spread of the paired lasers (10 and 11 cm for ODFW and DFG, respectively) were not counted.

DFG classified substrate in the video film as rock, boulder, cobble, or sand using categories simplified from Greene et al. (1999). A substrate layer was considered to be continuous until there was a break of ≥ 2 m or the substrate comprised less than 20% of total substrates for a distance of at least 3 m. Substrates were then combined into three habitat types: 1) hard (only rock and/or boulder), 2) mixed (rock and/or boulder with cobble and/or sand), and 3) soft (cobble and/or sand). ODFW used similar substrate classifications but included gravel and cobble in the hard and mixed habitat instead of in soft habitat. To make the data comparable, the amount of sand and gravel at Siletz Reef (approximately 6%) was subtracted from the hard and mixed category and added to the soft category (Table 1).

Computation of Density and Power Analysis

Twelve species were identified as potential candidates for this study with densities exceeding 0.01 per 100 m^2 in at least half of the 12 study areas (Table 2). Seven species that occurred in at least 10 study areas and had minimum and maximum densities differing by at least 300% were selected for analysis. These species were selected because they were sufficiently widespread and abundant to be amenable to analysis.

For evaluation of sample unit size, subunits within a single track line were combined to create transects measuring 50, 100, 200, 400, and 800 m². Fifty square meters was chosen as the smallest unit because it was close to the sample size used in most scuba surveys of fish in the local area and was sufficiently large to be measured accurately with an ROV. First, fish counts along each track line were divided into 25-m² (approximately 3 by 8 m) transect subunits. Subunits with more than 50% soft-only habitat were excluded because the objective was to measure density of fishes on predominantly hard bottom substrate. In addition to subunits with 100% hard or mixed substrate, subunits with 50% to 100% hard or mixed substrate were used to allow inclusion of hard or mixed/sand interfaces. A starting point for concatenating subunits was then selected randomly (Figure 3). Subunits were combined into transects of appropriate size, excluding one subunit between each transect

		Densit	y (No.	
		per 10	$0 m^2$)	Number
Species	Common Name			of Study
		Min	Max	Areas
Chromis punctipinnis	blacksmith	0.02	2.54	8
Hexagrammos decagrammus	kelp greenling	0.01	0.57	6
Ophiodon elongatus ^a	lingcod	0.02	0.32	11
Oxyjulis californica	senorita	0.01	1.18	8
Rhacochilus vacca	pile perch	0.01	0.34	9
Sebastes carnatus ^a	gopher rockfish	0.01	0.10	10
S. caurinus ^a	copper rockfish	0.01	0.13	11
S. miniatus ^a	vermilion rockfish	0.04	0.51	11
S. mystinus ^a	blue rockfish	0.15	3.65	12
S. serranoides ^a	olive rockfish	0.02	0.33	10
S. serriceps	treefish	0.01	0.08	9
Semicossyphus pulcher ^a	California sheephead	0.01	0.17	10

Table 2. S	pecies	quantified	on at]	least 6	5 of the	12 study	areas	with	densities	exceeding
0.0	1 per									
100	$) m^2$									

^a Species that met selection criteria by occurring in ≥ 10 study areas with minimum and maximum densities differing by at least 300%

to avoid contiguous transects. Transects overlapping two lines were also discarded. This pattern was repeated for the entire length of the track line to ensure random transect placement (Figure 3). Because the number of transects and total sampling area decreased with increasing sample unit size (Table 3), all comparisons were made with equal sampling area.



Figure 3. Illustration of how (a) twenty-three 100 m^2 and (b) five 400 m^2 transects were created using (a) four and (b) sixteen 25 m^2 segments respectively. Shown are the same three track lines at the start of a hypothetical study area.

Density and standard deviation were regressed for each species and transect size at each of the study areas to calculate the corresponding regression equations and coefficients of determination (r^2). Variances were then calculated from these equations corresponding to a range of 300% increase in the mean (maximum density / 4) and used in the software Java Applets for Power and Sample Size: Two-sample t-test (pooled or Satterthwaite) (Lenth 2009) to calculate sample size needed for a t-test to be statistically significant.

Transect Size	Number of Transects	Hectares Sampled			
50	2,781	13.9			
100	1,600	16			
200	753	15.1			
400	299	12.0			
800	88	7.0			

Table 3. Number of transects and hectares sampled by transect size for all study areas combined.

A *t*-test was used for this power analysis because of its relative simplicity as a predictive tool. A two-tailed test was used for the evaluation of transect size and both oneand two-tailed tests were used to evaluate the area needed to sample. All tests were run with alpha = 0.05, power = 0.8, and unequal variances for values of 50%, 100%, 150%, 200%, and 300% differences in the mean. A lower limit of 50% was chosen because it was in the range of expected change for our program, although smaller effect sizes can be evaluated.

RESULTS

To illustrate changes in slope and other properties of the data, regressions between mean density and the standard deviation at each location for each transect size for lingcod and vermilion rockfish are shown in Figure 4. For all species, there was a strong relationship between mean density and the standard deviation for transect sizes of 50, 100, 200, and 400 m² (Figure 5). All coefficients of determination (r^2) exceeded 0.65 and most exceeded 0.80. For 800-m² transects, coefficients of determination exceeded 0.75 for four species (lingcod, gopher, blue, and olive rockfish) and were less than 0.45 for three species (copper and vermilion rockfish and California sheephead). The slope of the regression lines decreased with increasing transect size (Figure 6). For 800-m² transects, the slope was less than 1.0 for five of seven species and less than 0.4 for two species.

Smaller transect sizes were more efficient than larger transect sizes (Table 4). For instance, to detect an increase in density by a factor of 2.5 (150%), approximately 11% more sampling area was needed for 100-m² transects than for 50-m² transects (Table 5). Respectively, 22%, 28%, and 57% more area was needed for 200-, 400-, and 800-m² transects.

The amount of total sampling area needed to detect a significant difference in density increased with decreasing effect size (Table 6, Figure 7). Some species (e.g. copper rockfish) required more sampling effort than others (e.g. vermilion rockfish). To select the total



Figure 4. Linear regressions between mean density and standard deviation for each location and sample unit size for lingcod and vermilion rockfish.



Figure 5. Coefficients of determination (r^2) between mean density and the standard deviation for each species for all transect sizes.

amount of sampling area needed for the program, the minimum detectible effect size for the species of concern was calculated for a range of sampling areas (Table 6). For example, 3 ha was identified as the sample area required for a minimum detectible effect size of 150% for all species.



Figure 6. Slope of the regression lines between mean density and the standard deviation for each species for all transect sizes.



Figure 7. Minimum sampling area (ha) needed to detect a significant difference in mean density for effect sizes of 100%, 150%, 200%, and 300% for each transect size.

DISCUSSION

Because there was a strong relationship between average density and variance across a broad range of densities, it was possible to use power analysis to identify the optimal sample unit size and then predict the minimum detectible effect sizes for a given amount of sampling area. Overall, the smallest sample unit size (50 m²) was most efficient, but the difference between 50- and 100-m² transects was only 11%. On the other hand, 800-m² transects required 57% more sampling area than 50-m² transects. In addition, with 800-m² transects, regression coefficients for 3 of 7 species were sufficiently low that predicting the variance from the mean density was questionable.

The relationship between sample unit size and the mean and variance has been known for many years. Taylor (1953) determined that for many species sampled with a trawl, variance was approximately proportional to the mean. The distribution of catch per tow conformed to a negative binomial rather than a Poisson distribution because the fish were aggregated, not randomly distributed. He also concluded that with aggregated populations, smaller sample unit sizes are more efficient. Green (1979) stated that while many environmental biologists intuitively feel that a larger sample unit size is better, in fact, sample unit size does not matter with randomly distributed populations, and smaller sizes are better with aggregated populations.

Taylor (1953) and Green (1979) evaluated frequency distributions of sample values, not the distribution of organisms in space; however, spatial distribution (e.g. patch size) is also important. If a species has a regular patch size, then a particular sample unit size may be most efficient. However, in every case, we found the smallest unit to be most efficient.

This is most likely because habitat is variable, patch size is variable, and smaller transects break up aggregations. Since there were at most two sample units per line for 400 and one for 800 m² transects (see Figure 3), any gradient or large-scale patchiness in distribution will be reflected in the variance between sample units (lines). Smaller sample unit sizes disaggregate the patchiness.

Some authors (e.g. Aubry and Debouzie 2000; Dungan et al. 2002) have concluded that larger sample unit sizes are more efficient; however, they did not consider total sampling effort or potential spatial autocorrelation within longer transects. When sample unit sizes are compared relative to a given sampling effort (e.g. Schoenly et al. 2003; Kimura and Somerton 2006), smaller sample sizes are generally more efficient.

Once the sample unit size is selected, methods outlined in this paper can be used to compute the minimum detectable effect size for a given amount of sampling. With this information, the tradeoffs between being able to measure expected changes in species of interest and the cost of the program can be evaluated. Because the spatial distribution of species differs (i.e., some may be territorial and evenly spaced and others may school and be aggregated) the amount of sampling will differ among species. But, with the range of sampling effort needed for the species of interest in a simple table, the choices involved can be based on the data.

In summary, these results demonstrate that pre-existing data may be used to establish relationships between means and variances required for power analysis. They also demonstrate that, in general, smaller sample unit sizes are more efficient. With this type of analysis, it is now possible to design a sampling program *a priori* that at a known sampling cost will have sufficient power to measure changes in populations between study areas and treatments.

ACKNOWLEDGMENTS

Special thanks to A. Lauermann of Pacific States Marine Fisheries Commission (PSMFC) for help in the sampling design, field data collection, and analysis. Thanks to C. Pattison and M. Prall of DFG for field support and sampling design. Special thanks to S. Ahlgren, Y. Yokozawa, and E. Jacobsen of PSMFC for assistance with analysis and graphics in this paper. Special thanks are due to N. Kogut and M. Patyten of DFG, and E. Jacobsen for help in editing this manuscript. Thanks to Stan Allen of PSMFC for administrative and staffing support. Thanks to M. Amend, B. Stave, and A. Merems of Oregon Department of Fish and Wildlife (ODFW) for data collection, sampling design, and analysis. Thanks are due to D. Rosen of Marine Applied Research and Exploration (MARE) for participating in field operations and fundraising. We also thank S. Holtz of MARE and the Institute of Fisheries Research for contributing to field operations in southern and central California. C. Mobley, K. Peet, and S. Fangman from the Channel Islands National Marine Sanctuary provided vessel support. L. Moody and T. Shinn captained the R/V Shearwater, and E. Schnaubelt and P. York captained the F/V Marlene Rose and the R/V Elakha, respectively. Thanks to R. Kvitek of California State University, Monterey Bay, G. Cochrane of the United States Geological Survey, and T. Sullivan of Seavisual Consulting for seafloor mapping. Primary support was provided to DFG from the Sportfish Restoration Act with additional support from the Ocean Protection Council, The Nature Conservancy, and the North East Assistance Program. Support for ODFW was provided by Oregon Department of Land Conservation and Development through a grant from the NOAA office of Ocean and Coastal Resource Management. Finally, we wish to thank G. Stacey, M. Vojkovich, P. Wolf, P. Coulston, and J. Ugoretz for their support as DFG managers.

LITERATURE CITED

- Aubry, P. and D. Debouzie. 2000. Geostatistical estimation variance for the spatial mean in two-dimensional systematic sampling. Ecology 81:543-553.
- Carr, G. M. and A. Morin. 2002. Sampling variability and the design of bacterial abundance and production studies in aquatic environments. Canadian Journal of Fisheries and Aquatic Sciences 59:930-937.
- Barry, J. P. and C. H. Baxter. 1993. Survey design considerations for deep-sea benthic communities using ROVs. Marine Technology Society Journal 26(4): 20-26.
- Dungan, J. L., J. N. Perry, M. R. T. Dale, P. Legendre, S. Citron-Pousty, M.-J. Fontin, A. Jakomalska, M. Miriti, and M. S. Rosenberg. 2002. A balanced view of scale in spatial statistical analysis. Ecography 25:626-640.
- Elliott, J. M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Second Edition. Freshwater Biological Association Scientific Publication 25:159.
- Gibbs, J. P., S. Droege, and P. Eagle. 1998. Monitoring populations of plants and animals. BioScience 48:935-940.
- Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley and Sons, New York, New York.
- Greene, H. G., M. M. Yoklavich, R. M. Starr, V. M. O'Connell, W. W.Wakefield, D. E. Sullivan, J. E. McRea Jr., and G. M. Cailliet 1999. A classification scheme for deep sea habitats. Oceanologica Acta 22:663-678.
- Hayes, J. P. 1987. The positive approach to negative results in toxicology studies. Ecotoxicology and Environmental Safety 14:73-77.
- Karpov, K. A., A. Lauermann, M. Bergen, and M. Prall. 2006. Accuracy and precision of measurements of transect length and width made with a remotely operated vehicle. Marine Technology Society Journal 40:60-66.
- Kimura, D. K. and D. A. Somerton. 2006. Review of statistical aspects of survey sampling for marine fisheries. Reviews in Fisheries Science 14:245-283.
- Lenth, R. V. 2009. Java applets for power and sample size [Computer software]. Retrieved April 27, 2009 from http://www.stat.uiowa.edu/~rlenth/Power
- Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47:2-15.
- Schoenly, K. G., I. T. Domingo, and A. T. Darrion. 2003. Determining optimal quadrat size for invertebrate communities in agrobiodiversity studies: a case study from tropical irrigated rice. Environmental Entomology 32:929-938.
- Steidl, M., J. P. Hayes, and E. Schauber. 1997. Statistical power analysis in wildlife research. Journal of Wildlife Management 61:270-279.
- Toft, C. A. and P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. American Naturalist 122:618-625.
- Taylor, C. C. 1953. Nature of variability in trawl catches. Fishery Bulletin 54:145-166.
- Wakefield, W. W. and A. Genin 1987. The use of a Canadian (perspective) grid in deep-sea photography. Deep-Sea Research 34:469-478.

Received: 15 May 2009

Accepted: 9 October 2009

Associate Editor: T. Barnes